From Screening to Intervention: Instructional Planning for Students Who Struggle with Whole Number Computation

A Dissertation
SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA
BY

Nicole Marcella McKevett

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Robin Codding, Dr. Amanda Sullivan

May 2021

Acknowledgements

First and foremost, I would like to thank my husband, Bob McKevett, for his unwavering support, patience, and kindness throughout my educational career. I would not have been able to do this without his support. He has been a sounding board, has always pushed me forward, and pulled back the reigns when I needed to slow down. This dissertation has my name on it, but it is truly the product of both of our hard work.

My sincerest gratitude goes to my advisor, Dr. Robin Codding. With you, my love for data-based decision making in education found a home, and I am grateful for your guidance over the past years. To my co-advisor, Dr. Amanda Sullivan, thank you for your leadership in the UMN school psychology program, and the field at large. Your kindness and advice has reminded me of what is important and how to prioritize the things we value in life. Many thanks to Dr. Kristen McMaster, for everything you have done for me in the past years of working together; I learned so much from you about ethical and effective research practices, what a fantastic mentor to graduate students looks like, and how to keep the love of learning and discovery at the front of what we do. Finally, thank you to Dr. Kimberly Gibbons, for believing in my potential as a systems-change agent, a leader in the field, and bringing me to CAREI. In working with you, I have felt appreciated and have learned to believe in my own knowledge and skills.

Last, but most certainly not least, I must thank my parents, Jonathan and Pamela Weber. There was certainly a method to your madness that has led to all three of your children pursuing a doctoral degree: a DPT, PhD, and MD. You have so much to be proud of, and for that I thank you, from the bottom of my heart.

Dedication

Right now, the world is coming to the end of 2020, a year that has been challenged by the COVID-19 pandemic. As of December 15th, 2020, nearly 300,000 lives have been lost as a result of COVID-19 in the United States. If I were to read all of the names of those that have been lost to this day, it would take me 10 days. This dissertation is dedicated to all those who have been lost as a result of this pandemic, to the health care professionals and essential workers that have kept our country afloat, and to the educators who are experiencing unprecedented challenges in continuing teaching throughout the COVID-19 pandemic.

Abstract

This dissertation conducted two studies that examined two methods of instructional planning to effectively match students to a whole number computation intervention that would best meet their needs. Study 1 was a systematic synthesis of all studies that used brief experimental analysis (BEA) to determine an effective mathematics intervention for students. Sixteen studies that included 67 participants and used a BEA to identify the most effective mathematics intervention were located. Results of Study 1 indicated that the majority of BEAs compared skill and performance interventions on computational fluency; however, the methodology across the included studies greatly varied. The second study evaluated a gated screening framework that included STAR Math, AIMSweb™ Math Computation (MCOMP), and a can't do/won't do assessment using AIMSweb™ Subskill Mastery Measure-Addition/Subtraction (SSMM-Add/Sub). A standard BEA was used to evaluate which of two interventions, modeling with error correction or explicit timing with reward, was most effective for each student. Analyses determined whether each of the screening measures could accurately differentiate between the students who benefitted the most from each intervention and accurately predict the outcomes of the BEA. Statistically significant differences were yielded for SSMM-rate but not STAR Math or MCOMP. STAR Math and SSMM-rate were able to predict which intervention was most effective. A cut score analysis indicated that the optimal cut score for SSMM-rate to differentiate between interventions was 13 DCPM.

Keywords: whole number proficiency, brief experimental analysis, gated screening

Table of Contents

List of Tables

List of Figures

Study 1 Figures

Chapter 1

Proficiency in math is an important factor for individual citizens and nations as a whole, as it is related to college and career readiness, career options, and future income potential (Geary, 2004). Specifically, proficiency in math skills up to Algebra II is positively correlated with access to and graduation from college and earning in the top quartile of income (NMAP, 2008). However, on average, U.S. students have not yet achieved math proficiency at a level to compete well against other industrialized nations (National Assessment of Educational progress [NAEP], 2019; NMAP, 2008). On the Trends in International Mathematics and Science Study (TIMMS), fourth grade U.S. students scored below 14 other industrialized nations and eighth grade students scored below 10 other industrialized nations (Mullis et al., 2012). Additionally, 41% of fourth graders and 34% of eighth graders scored at or above proficiency in mathematics on the 2019 NAEP. There also were insignificant changes in performance from 2017 to 2019, with average proficiency level remaining the same in eighth grade and increasing by just 1% in fourth grade (NAEP, 2019).

Efforts to Improve Proficiency in Mathematics

While a concern regarding math achievement in the U.S. is the overall proficiency level, an additional concern is that students with lower math achievement levels also experience slower growth than the general population of students, particularly in elementary school (Wei et al., 2012). On the 2019 NAEP, the average math scores of those who scored at the 10th and 25th percentiles was lower than the average score at these percentiles in 2017 and in 2009, suggesting that there are not only persistent, but widening gaps in achievement growth for students who are performing below average

(NAEP, 2019). For those students who demonstrate math difficulties and slower growth rates, deficits in whole number proficiency are often persistent (Geary, 2004). Whole number proficiency is an essential skill for overall mathematics proficiency. Typically defined as the efficient and accurate computation of math calculation in addition, subtraction, multiplication, and division, whole number proficiency is an indicator for the development of more complex math skills (e.g., solving complex problems and interpreting abstract mathematical concepts; Patton et al., 1997), a predictor of outcomes on state assessments (Shapiro et al., 2006), and important for the development of independent living skills (Patton et al., 1997).

Multi-Tiered System of Supports in Mathematics

One way to provide supplemental intervention for students who struggle to master foundational skills is to implement a multi-tiered system of support (MTSS; Jimerson et al., 2015). MTSS is a comprehensive school improvement framework, that emphasizes high-quality core instruction and uses resources to implement a Response to Intervention (RtI) process with students who demonstrate academic and behavioral difficulties (Jimerson et al., 2015). In MTSS frameworks to improve math achievement, core instruction is designed to keep average and above-average achieving students on track, while also differentiating instruction to meet the needs of those with math difficulties (Doabler et al., 2012). For those who do not make expected progress with differentiated core instruction, an RtI process is used to systematically identify student needs, provide targeted interventions, and monitor student growth (Jimerson et al., 2015). By providing appropriately matched supplemental intervention, educators have been able to improve the math performance of elementary students who have persistent difficulties with basic math skills (Dennis, 2015).

While the majority of research on RtI has been conducted in the area of reading, in the last decade, increasing attention has been paid to how to implement RtI in mathematics (Lembke et al., 2012). One recommended approach is to use the problem-solving model, which includes five key steps: (1) problem identification, (2) problem analysis, (3) plan development, (4) plan implementation, and (5) plan evaluation (Deno, 2015; Fuchs & Fuchs, 2006; Marston et al., 2003). At the problem identification step, universal screening procedures are used to evaluate the quality of core instruction and identify which students are achieving below grade-level expectations (Ardoin et al., 2005). At the problem analysis step, educators can engage in the process of instructional planning, to determine what skills to teach and how to teach (Zigmond & Miller, 1986). The problem analysis step provides information on the instructional target and hypotheses about which instructional procedures will be most appropriate for a student. The procedures used at the problem identification and problem analysis steps are critical to matching students to an appropriate intervention, so as not to exhaust educator time and school resources (Cook et al., 2018). It is also critical to the validity of RtI, because inappropriate intervention selection can actually increase the number of students who do not respond to intervention, which may lead educators to feel that the RtI process is not effective (VanDerHeyden et al., 2005).

Universal Screening in Mathematics. The delivery of an effective RtI approach in mathematics depends on the use of accurate and efficient universal screening procedures and measures (Foegen et al., 2007). The primary purpose of conducting universal screening is to identify students who are at risk for not meeting end-of-year grade-level standards (Johnson et al., 2010). After students are identified as at-risk, they are typically routed to receive supplemental intervention (Jenkins et al., 2014). However, for supplemental interventions to work effectively, the measures used for universal screening must be reliable and valid, result in a high rate of true positives (i.e., the students who are identified as at-risk are truly at-risk) and be able to accurately inform the process of routing students to interventions (Compton et al., 2010; Jenkins et al., 2014). To enhance the accuracy of routing students to the right supplemental intervention, educators must use more sophisticated decision-making processes, such as problem analysis, rather than relying on resource and time intensive, trial-and-error methods of intervention selection (Christ & Aranas, 2008; VanDerHeyden, 2013). Doing so will allow educators to answer the questions "what works, for whom, and under what conditions," which is essential for guiding effective service delivery in a timely manner, thereby preventing widening achievement gaps (Miller et al., 2020).

Gated Screening in Mathematics. One approach for using more sophisticated decision-making processes for the purpose of routing students to an effective supplemental intervention is the use of a gated screening framework (Compton et al., 2010). A gated screening framework uses multiple measures to improve the identification of students who are at risk (Compton et al., 2010). A gated screening framework involves delivering follow-up screening measures, after a student has been identified as at-risk on a strong, initial universal screening measure (Van Norman et al., 2018). This approach, while it does involve using multiple measures, has been identified as one of the most time- and cost-effective approaches in applied, educational settings (Van Norman et al., 2018). In reading, using multiple measures is not necessary, as schools can use a single measure of oral reading fluency (CBM-R) across several grades to identify students who are at risk and for instructional planning (Szadokierski et al., 2017). However, in mathematics, a gated screening process may be necessary, as the measures that operate best at different grade levels to identify which students are at risk for not meeting end-of-year grade-level standards varies and little is known about how to use these measures for instructional planning (VanDerHeyden et al., 2017; Van Norman et al., 2018).

Gated screening frameworks assume that collecting additional data will provide unique information about a student's performance, over and beyond that which was captured with the initial measure (Van Norman et al., 2018). However, one caution in this approach is the use of two measures that are highly correlated with each other, as doing so will provide redundant information as the first measure (VanDerHeyden, 2013; Van Norman et al., 2016). Gated screening frameworks have primarily been identified as an approach to improving the identification of students who are at-risk; however, some researchers have also noted the potential of using this framework for gathering instructionally relevant information and identifying students who are a good fit for intervention services that are available at the school (Van Norman et al., 2018). In math, recommendations for this process include the use of a strong initial screening measure, to accurately identify students who are at-risk. Then, educators can use a criterion-referenced or subskill mastery measure, such as a brief fluency assessment (Vaughn & Fletcher, 2012), along with theories of instructional planning, to match students to an intervention.

Instructional Planning in Mathematics

Different theories of instructional planning can be used to adequately conduct problem analysis, for the purpose of matching a student to an appropriate intervention (Cook et al., 2018). This includes the instructional hierarchy (Haring & Eaton, 1978) and the use of functional analysis (Daly et al., 1997).

Instructional Hierarchy

The instructional hierarchy (Haring & Eaton, 1978) describes how students' skills develop through four main stages: acquisition, proficiency, generalization, and adaption. Haring and Eaton (1978) described how students' skills change as they progress through this hierarchy, as do the instructional procedures that are most likely to be effective for students whose skills are at each stage. For example, a student at the acquisition stage typically performs a skill with less than 90% accuracy. Once the student reaches 90% accuracy, they transition to the fluency stage, where activities such as timed, repeated practice are likely to help them to improve their accuracy and fluency.

In math, the instructional hierarchy was primarily described with respect to the basic computational processes of addition, subtraction, multiplication, and division (Lovitt, 1978). Lovitt (1978) describes how a child who is first learning computation may perform a skill with 40% to 60% accuracy. However, as they begin to acquire the skill and move onto the stage of fluency, they perform the computation problems with close to 100% accuracy, but below a desired rate of fluency. For students who exhibit computation skills within the acquisition stage, instructional strategies that include modeling are typically most effective (Lovitt, 1978). For students who exhibit computation skills within the fluency stage, instructional strategies that include timing and contingent reinforcement are typically most effective (Lovitt, 1978). Two studies in math have demonstrated the connection between the instructional hierarchy and effective instructional procedures. Codding et al. (2007) found that students who are within the acquisition stage of whole number proficiency typically best respond to modeling and error correction strategies, while students who are within the fluency stage of learning typically benefit most from repeated and timed practices. Additionally, a meta-analysis on the effectiveness of acquisition and fluency interventions found that acquisition interventions resulted in larger effect sizes among children with acquisition level skills, but only moderate effects for students with fluency level skills (Burns et al., 2010).

Functional Analysis

In 1997, Daly et al. proposed a framework for applying functional analysis to academic skills, to help educators determine how to teach. Daly et al. (1997) described how taking a functional approach to understanding academic difficulties is useful, because these factors allow for direct manipulation and led themselves to instructional planning. To conduct a functional analysis, Daly et al. (1997) proposed that educators (a) generate hypotheses about the possible ways in which a student can fail and (b) expose the student to brief test conditions that mimic the hypotheses. This approach is consistent with recommendations from Ysseldyke & Alogozzine (1984), who also recommended conducting trial teaching assessments and evaluating the effectiveness of instruction, to determine which instructional approach will be most responsive to the needs of a particular child.

To conduct the first step of generating hypotheses about why a student is struggling, Daly et al. (1997) recommended considering five common hypotheses for why students fail include: (a) they do not want to do it, (b) they have not spent enough time doing it, (c) they have not had enough help doing it, (d) they have not had to do it this way before, and (e) it is too hard (Daly et al., 1997). However, educators might also consider the instructional hierarchy or can't do/won't do theoretical framework as well (Burns et al., 2010; Haring et al., 1978). A can't do/won't do assessment can be used to determine if a student's academic difficulties are due to a skill (can't do) or performance (won't do) deficit (VanDerHeyden & Witt, 2007). A can't do/won't do assessment has also been used within a screening model in math to verify if students who are identified as at-risk were an appropriate match for the skill-based interventions that were available in the school (Ardoin et al., 2005).

In academics, there are several considerations that must be made when conducting a functional analysis in academics. First, instructional strategies that are used have to be brief, easy to implement and be able to produce immediate and noticeable improvements in performance. Second, a multielement design, such as a brief experimental analysis (BEA), can be used to establish a baseline level of performance, followed by alternately repeating the intervention conditions (Daly et al., 1997). Third, there are several key considerations related to the assessment materials. The assessment materials must: (a) be sensitive to short-term gain, (b) be of equal difficulty across conditions to assure that differences in outcomes are not due to differences in difficulty of the outcome measures, and (c) the assessment materials should have considerable overlap with the intervention materials (Daly et al., 1997). BEAs have been used to conduct functional analyses in the areas of reading (e.g., Jones & Wickstrom, 2002), writing (Burns & Wagner, 2008), and math (Mong & Mong, 2012). With respect to BEAs, much more is known about how to conduct a BEA in reading and how to use this information to guide instructional planning for students who are struggling; however, much less is known about how to use these procedures in the area of math, such as what intervention conditions would be appropriate to test within the BEA and how best to measure intervention outcomes.

Purpose

This study will evaluate two methods for instructional planning in mathematics: BEAs and a gated screening framework. Study 1 will be a systematic literature review on the use of BEAs in mathematics. In reading, there are clear recommendations for how researchers and practitioners should complete a BEA; however, in math, much less is known. Conducting a systematic synthesis of studies that have implemented BEAs in math will be helpful in providing recommendations for future research and for practitioners, who may be interested in using this approach to match students to appropriate supplemental interventions. Study 1 sought to answer the following research questions:

1. What are the characteristics of participants and settings (e.g., gender, grade, race, urbanicity) of studies that have implemented BEAs in math?
2. What methodology characteristics are most commonly used in the math BEA literature (e.g., purpose and type of BEA, experimental design, primary dependent and independent variables, type of primary outcome measure)?
3. What are the outcomes for students in studies that implement BEAs in math (i.e., which interventions were found to be most effective)?

The primary purpose of Study 2 was to evaluate a gated screening framework in math that could be used for instructional planning for students who are at-risk in whole number proficiency. STAR Math and AIMSweb™ Math Computation (MCOMP), two universal screening measures, were used to identify students who were at risk. Similar to the methods of Ardoin et al. (2005), a can't do/won't do assessment, using the AIMSweb™ Subskill Mastery Measure Add/Sub (SSMM-Add/Sub), was used to verify if the students that were identified as needing a supplemental intervention in whole number proficiency were an appropriate match for the skill-based interventions that were available. As part of instructional planning, a BEA was implemented to verify which intervention was most effective for each student (modeling with error correction or explicit timing with reward). Then, it was determined if each of the included screening measures could be used to accurately differentiate between the students who benefitted the most from each intervention and if each measure could accurately predict the outcomes of the BEA. This was done to determine if educators could use these screening measures to accurately assign students to an intervention without needing to use the time and resources needed to engage in a BEA. The research questions guiding this study were:

1. Are there statistically significant differences in students' performance on the measures included in the gated screening framework (STAR Math, MCOMP, SSMM-Add/Sub), between those for whom modeling with error correction was most effective and those for whom explicit timing with reward was most effective?
2. Which variables in the gated screening framework (STAR Math, MCOMP, can't do/won't do assessment, and SSMM-Add/Sub) best predicted which intervention was the most effective in the BEA?

The current study also addressed an exploratory research question:

3. What are the cut scores on the variables that are significantly different between each effective intervention group, to provide enough specificity and sensitivity to predict which intervention a student will respond best to?

Chapter 2

Brief Experimental Analysis of Math Interventions: A Synthesis of Evidence

Many children in the United States struggle to demonstrate competence in math. This is indicated by multiple metrics of international math achievement (Trends in International Mathematics and Science Study [TIMMS]; Program for International Student Achievement [PISA]), which show that U.S. students' average math score is lower than the average score of peers from other developed countries (National Center for Education Statistics [NCES], 2015; Mullis et al., 2012). Additionally, on a recent assessment of national performance, 40% of fourth graders and 34% of eighth graders were at or above proficiency in math (National Assessment of Educational Progress [NAEP], 2017). These data are concerning, given there was little change in the nation's performance at these grade levels from 2015 (NAEP, 2017). Moreover, performance for students falling below the 25th percentile declined from 2015 to 2017. Low math achievement has been associated with low personal and employment outcomes for both men and women, such as lower rates of full-time employment, lower salaries, and less potential for position advancement (Geary, 2011). Taken together, this suggests a need for U.S. educators to not only enhance core math instruction, but to also provide remediation before achievement discrepancies become too large to remediate, using data-based decision making to inform math instructional decisions (VanDerHeyden & Burns, 2005).

MTSS and the Problem Solving Model

School systems have made efforts to enhance core and supplemental math instruction by implementing a multi-tiered system of supports (MTSS). MTSS is a comprehensive framework of resource allocation that includes several key components: (a) high quality core instruction, (b) universal screening, (c) evidence-based supplemental intervention, and (d) data-based decision making (Burns et al., 2015). MTSS has been characterized in two ways: a standard protocol approach, where all children who do not meet a specified criterion receive a common intervention (Fuchs et al., 2003), and through the application of a problem-solving model. Implementation in reading is the main area for which MTSS has been implemented as indicated by a 2010 survey in which 90% of elementary schools surveyed reported some degree of implementation (from investigation to full implementation; Spectrum K-12, 2010). Implementation of MTSS in math follows reading as the most predominant area in elementary, middle, and secondary schools (Spectrum K-12, 2010).

The problem solving model that is embedded within an MTSS framework includes five steps: (a) identify the problem to be solved, (b) define the problem, (c) explore alternative solutions, (d) apply the selected intervention, and (e) evaluate the effects (Deno, 2015). First conceptualized as a way to individualize educational plans for students with significant learning or behavioral difficulties (i.e., data-based program modification; Deno & Mirkin, 1977), this model can also be applied to develop intervention plans for students at all levels of an MTSS system (Deno, 2015). School psychologists have a unique skill set to be valuable leaders in applying the problem solving process for students at Tiers 2 and 3 of the MTSS framework (Burns & Coolong-Chaffin, 2016). Specifically, school psychologists have an important role in collecting and interpreting assessment data and identifying evidence-based interventions. By connecting knowledge about what interventions to use and how to implement them, school psychologists can fill a research-to-practice gap in the implementation of evidence-based interventions (Burns & Coolong-Chaffin, 2016; Forman et al., 2013).

Brief Experimental Analysis

Brief experimental analysis (hereafter referred to as BEA) is a set of single-case designs used to evaluate variables that have an immediate impact on a specific target behavior or skill (Daly et al., 1997; Gast & Ledford, 2014; Jones & Wickstrom, 2002; McComas et al., 1996). By helping to identify both an effective and *efficient* strategy, a BEA can be used to implement the problem solving model, ensuring that precious resources of an MTSS system, such as time and people, are not misapplied (Deno, 2015; Jones & Wickstrom, 2002). It has been suggested that the use of BEAs to predict effective interventions may be one of the most promising approaches for struggling students (McComas & Burns, 2009). Additionally, considering the fact that conducting a BEA can take approximately the same amount of time as a standardized assessment, it has been suggested that the use of strategies like a BEA can transform the role of school psychologists from diagnosing special education eligibility to finding out what works for students (Jones & Wickstrom, 2002; Martens & Gertz, 2009).

Originating from the theoretical foundations of applied behavior analysis and single-case design, BEAs involve repeated measurement of a dependent variable over time (progress monitoring), the implementation of pre-specified interventions, and replication of intervention effects (Martens & Gertz, 2009). BEAs fall in the class of multi-element design, which includes a baseline period, and the comparison of two or more treatments (Barlow & Hayes, 1979; Daly et al., 1997). Once all conditions have been presented, a mini-reversal is conducted, where the condition that produced the greatest increase in performance is repeated just after baseline or a condition that produced a lower increase (Daly et al., 1997; McComas et al., 1996). A standard BEA design includes a minimum of three sessions per condition; however, there are two alternative designs: (a) an abridged BEA design (e.g., Jones & Wickstrom, 2002) which requires a single session per test condition (Daly et al., 1997); and (b) a one-trial BEA design which includes a single reward or reward plus instruction trial (Anderson et al., 2013).

There are three variations of BEAs that have been conceptualized and used in educational research (Andersen et al., 2013). The first variation includes a baseline session followed by the comparison of separate treatment packages. For example, Jones and Wickstrom (2002) compared the effects of reward, repeated reading, phrase drill, and easier reading material on students' oral reading fluency. The second variation includes systematic comparison of interventions from least intensive (time to complete and amount of adult assistance required) to most intensive (McComas et al., 1996). For example, Daly et al. (1999) systematically ordered the presentation of interventions designed to improve oral reading fluency from least intensive to most intensive and compared student performance in each phase to baseline. In their interpretation of the results, the authors considered not only the impact of each intervention package on student performance, but also the amount of time each phase took to implement, and the amount of adult assistance required (Daly et al., 1999). The final variation involves determining whether a student's difficulties are due to a skill or performance deficit (Lentz, 1988). Researchers first collect baseline performance data, then provide contingent reinforcement if the student can improve their baseline performance by a pre-specified criterion (e.g., Duhon et al., 2004). If a student meets or exceeds the pre-specified criterion, it is hypothesized that the student has a performance deficit. If they do not meet the pre-specified criterion, it is hypothesized the student has a skill deficit (Duhon et al., 2004). Researchers then implement an instructional or motivational intervention to validate their hypothesis (Lentz, 1988).

To determine which intervention in the BEA is most effective compared to baseline, many researchers use a percent criterion (e.g., 20% criterion, 50% criterion). Noell et al. (2001) were the first to propose the use of the 20% criterion. Based on the recommendations of Carnine et al. (1990) that an increase of 40% would be an appropriate goal for students who receive an oral reading fluency intervention, Noell et al. (2001) suggested that a 20% increase would be an appropriate goal after a brief presentation of an intervention. However, Daly et al. (1999) indicated that pre-determined increase approaches may be problematic depending on the student's baseline level. For example, increasing performance by 50% may be more likely for a student whose baseline is 10 digits correct per minute (DCPM) than for a student whose baseline is 30 DCPM. Given the idiographic nature of BEAs, it may be important to consider the combination of the 20% increase criterion, which has been the most used percent criterion in academic BEAs, and methods that consider individual differences in responding (e.g., visual analysis or mean performance difference; Daly et al. 1999).

Another important consideration when interpreting the outcomes of BEAs is the degree to which assessment materials and instructional content overlap (Daly et al., 1996). When assessment materials have fewer opportunities for students to demonstrate the target skill from the intervention session, the assessment results are likely to underestimate the impact of the intervention (Daly et al., 1996). It is also essential that the assessment materials are equally difficult, so that outcomes do not reflect differences in the difficulty level of outcome measures, as well as different from each other, to reduce carryover effects (Daly et al., 1997).

The utilization of a theoretical framework to select interventions is an important aspect of using a BEA to determine the most effective intervention. In seminal research on the use of functional analysis applied to academic skills, Daly et al. (1997) proposed one such framework, within which five common factors that can affect student academic performance are outlined: (a) they do not want to do it, (b) they have not spent enough time doing it, (c) they have not had enough help to do it, (d) they have not had to do it that way before, or (e) it is too hard. The instructional hierarchy is another framework for conceptualizing academic difficulties (Burns, et al., 2010; Daly et al., 1997; Haring et al., 1978). The instructional hierarchy includes five main stages of academic skill development: acquisition, fluency, maintenance, generalization, and adaptation (Haring et al., 1978). The majority of BEAs that focus on academic skills have implemented intervention packages that target acquisition and fluency deficits, while the primary dependent measure of these studies has been a fluency outcome (e.g., WCPM, DCPM; Codding et al., 2009; Jones & Wickstrom, 2002).

Applications of BEAs in Academics

In educational research, BEA methodology has been used in the areas of reading (e.g., Eckert et al., 2002; Jones & Wickstrom, 2002), writing (e.g., Burns et al., 2009), and math (e.g., Mong & Mong, 2012; Reisener et al., 2016); however, a majority of the research that has applied BEA methodology to academic skills has examined intervention effects on oral reading fluency. Enough research has been conducted in this area that researchers have been able to apply meta-analytic techniques to determine what types of effects are required within a BEA to determine the most effective intervention in oral reading fluency (Burns & Wagner, 2008). Recently, Burns et al. (2017) examined the effect of standard error of measurement (SEM) on intervention decisions across all reading BEA studies, finding that only one in four intervention comparisons within the BEA resulted in a performance difference greater than the SEM.

In reading BEAs, researchers typically implement a consistent suite of interventions, which allows for syntheses like those conducted by Burns and Wagner (2008) and Burns et al. (2017). This suite includes contingent reward, performance feedback, listening passage preview, repeated reading, and phrase drill (Burns & Wagner, 2008). Each of these interventions is designed to target a specific function of the students' low performance. Contingent reward targets low motivation; repeated reading targets a lack of practice; performance feedback and phrase drill target a lack of instructional feedback; and error correction and listening passage preview target a lack of modeling (Burns & Wagner, 2008). However, far less is known about math BEAs, particularly with respect to what interventions should be implemented and compared, what outcome variable to measure and how to measure it, and how to determine the most effective math intervention.

Purpose

BEA has been identified as a method of bridging the research-to-practice gap of evidence-based interventions within a problem solving MTSS framework. School psychologists have a unique skill set to connect assessment outcomes to intervention implementation and can use efficient single-case design procedures to determine what works and with whom. In the reading BEA literature, there are clear recommendations of what interventions should be implemented within a BEA to target hypothesized functions. However, much less is known in research and practice about how to implement BEAs in math. Clearly defining these research and practice guidelines is essential in guiding researchers and practitioners toward filling a gap in the implementation of MTSS in math. Therefore, the purpose of the current study is to synthesize the research on math BEAs, describe the strengths and weaknesses of this line of research, and provide recommendations to both researchers and practitioners. The current study sought to answer the following research questions:

1. What are the characteristics of participants and settings (e.g., gender, grade, race, urbanicity) of studies that have implemented BEAs in math?
2. What methodology characteristics are most commonly used in the math BEA literature (e.g., purpose and type of BEA, experimental design, primary dependent and independent variables, type of primary outcome measure)?
3. What are the outcomes for students in studies that implement BEAs in math (i.e., which interventions were found to be most effective)?

Method

Literature Search and Inclusion Criteria

A search of studies that implemented math BEAs was conducted ending in September 2019. First, an electronic search using the databases EBSCOhost, PsychINFO, ProQuest-Dissertations, and Web of Science was completed. The following search terms and roots were used: math* and "brief experimental analysis" (*n* = 239), computation and “brief experimental analysis” (*n* = 0), arithmetic and “brief experimental analysis” (*n* = 0), calculation and “brief experimental analysis” (*n* = 0), “brief experimental analysis” (*n* = 130), and math* and BEA (*n* = 22). Next, the reference sections of included studies were searched to identify additional manuscripts or dissertations that were not located in the electronic search (*n* = 3). Finally, relevant journals (*Assessment for Effective Intervention, Journal of Applied Behavior Analysis, Journal of Applied School Psychology, Journal of Behavioral Education, Psychology in the Schools, School Psychology Review,* and *School Psychology Quarterly*) were hand searched. The primary search yielded a total of 394 articles. After the removal of duplicate articles (*n* = 95), 299 studies were reviewed for inclusion in the study.

Five inclusion criteria were applied. First, the study had to implement an intervention intended to improve student performance on a math outcome. Studies were excluded if a primary dependent variable of the study was not a math outcome (e.g., digits correct per minute, total number of problems correct, etc.). Second, the study had to be an experimental study that used brief experimental analysis as its primary methodology. A brief experimental analysis was defined as the alternating of two or more independent variables within a multi-element design (McComas et al., 1996). Third, the study was required to be a dissertation or published in a peer-reviewed journal; master’s theses were excluded. Fourth, the study sample had to consist of school-age students (kindergarten through 12th grade). Finally, a copy of the study translated to English had to be accessible.

Two hundred and ninety-eight studies were included through the original search and 284 were excluded. Of the studies that were excluded, 183 did not implement an intervention intended to improve math outcomes, 22 were not experimental, and 79 did not use brief experimental analysis as their primary methodology. Title and abstract review resulted in 19 studies identified for potential inclusion. Next, the full-text versions of these studies were reviewed, and five more studies were excluded because the studies did not include math as a primary outcome variable or BEA as the primary methodology. The primary search resulted in a total of 15 relevant manuscripts or dissertations. One manuscript (Reisener et al., 2016), included two studies, which resulted in a total of 16 included studies.

Coding Procedure

The primary author developed a coding sheet to extract relevant information in the following categories: (a) participants, (b) setting, (c) BEA methodology, and (d) extended analysis methodology. Each study was independently coded by the first author, with 30% of the studies coded for inter-rater agreement (described below).

Participant Characteristics

To gather information about the students included in each study, we coded the total number of participants, gender, grade level, special education eligibility (identified by the school or researchers), race (Asian, Black, Hispanic and/or Latino, Native American and/or Alaskan, Pacific Islander and/or Hawaiian, White, and Multiracial), and English Language Learner status.

Setting Characteristics

To describe the setting of each study, we coded the geographic location of the study, the percent of free or reduced lunch at the school, and the urbanicity of the school (urban, suburban, rural, or other).

BEA Characteristics

The BEA methodology used at the study level was coded, in addition to the characteristics of the BEA that varied at the individual student level.

Study-Level BEA Characteristics. First, the purpose for conducting the BEA, defined by the researchers, was coded (i.e., skill vs performance functional analysis, best skill intervention, best performance intervention). Then, the type of BEA that was used (standard or abridged) was coded. A standard BEA was defined as a multielement design, where multiple, alternating sessions using each intervention were conducted (Daly et al., 1997). An abridged BEA was defined as a multielement design that presented each intervention condition only once (Jones & Wickstrom, 2002; McComas et al., 1996). The experimental design of the extended analysis was also coded (e.g., multielement, multiple baseline).

Next, we coded the criteria used to determine (a) if an intervention was more effective than baseline and (b) if an intervention was more effective than another intervention within the context of the BEA. Specifically, we identified what criteria the researchers used to determine if an intervention was more effective than baseline, in addition to the criteria used to determine which intervention was the most effective within the BEA (e.g., visual analysis, a 20% criterion). Visual analysis was defined as viewing student performance graphs and visually determining if there was a level change between conditions, indicating the relative strength of the intervention on the dependent variable (Gast & Ledford, 2014). The 20% criterion was defined as calculating if the student response to an intervention was 20% higher than baseline performance (Noell et al., 2001). Other methods used to determine intervention effectiveness (e.g., 50% criterion) were coded and qualitatively described.

We next coded the primary dependent and independent variables included within the BEA. We also coded whether the screening measures used to identify study participants were researcher-developed or standardized. Standardized measures were defined as norm-referenced with formal procedures for administration and scoring (Bond, 1996; e.g., AIMSweb MCOMP, the Iowa Test of Basic Skills, and NWEA MAP). Researcher-developed measures were defined as screening measures made by the researchers specifically for the study.

Next, we coded specifics about the outcome measure used to capture student performance on the primary dependent variable during the BEA. Specifically, we coded (a) whether the primary outcome measure was researcher-developed or standardized; (b) what method the researchers used to develop their measure (e.g., randomized problems, Math Worksheet Generator from interventioncentral.org, probe sets); (c) whether researchers assigned problems to treatment conditions within the BEA or assigned problems to specific worksheets; and (d) whether the researchers ensured equivalency of the outcome measures as well as considered SEM when interpreting intervention effects across phases.

Student-Level BEA Characteristics. For each student within each study, we visually analyzed each graph, examining the level changes across conditions, and determined (a) the hypothesized deficit from a skill vs performance assessment, (b) the best intervention in the BEA, (c) if the results of the extended analysis matched the BEA, and (d) whether the results of the extended analysis were differentiated. Extended analysis results were defined as differentiated if there was a consistent difference in level, trend or variability between the intervention phases, allowing for a visual analyst to determine which was the most effective intervention (Backman et al., 1997).

Quality Indicators of Included Studies

Quality indicators from Xin (2008), which are based on the single case design criteria outlined by Horner et al. (2005) and developed to provide qualitative markers of high-quality single-case design methodology, were adapted for the purposes of this study. Adaptations to this criteria included modifying the number of data points required (as an abridged BEA naturally has fewer than three data points per condition), elimination of specific baseline criteria (as a BEA does not require the inclusion of a baseline with three or more data points that are stable prior to beginning intervention), and the addition of specific criteria on the extended analysis that were in alignment with the *What Works Clearinghouse Reviewer Guidance* for alternating treatments design. Variables that were specific to the use of study-level BEA methodology were used to determine the overall quality of the literature included in this systematic review.

Inter-rater Agreement

Three school psychology graduate students (one PhD level, two specialist level) received a one-hour training session from the primary author on conducting inter-rater agreement. First, the inclusionary criteria and the exclusion codes were defined. Next, the coding sheet was explained, along with an explicit definition of each code. The primary author also reviewed the basics of BEA methodology and visual analysis with the graduate students. Finally, an included study was shown to the students. The group discussed this study and practiced coding the article, with immediate assistance and feedback from the primary author. The three graduate students were then each assigned one-third of the studies located in the primary search (~100 studies each), in addition to one or two studies to code (five total studies). The coders independently applied the inclusion criteria to all studies and independently coded the assigned studies. The coders then met with the primary author to discuss any discrepancies in coding. Differences were resolved by reexamining the studies to settle on the most appropriate code. The inter-rater agreement for inclusion criteria and the individual studies was 95%.

Data Analysis

Descriptive analyses were conducted to provide quantitative information about the various study characteristics and student outcomes that were coded, in addition to the overall quality of the literature included in this review.

Results

Participant Characteristics

Table 1 provides a summary of participant and setting characteristics. Studies included a total of 67 students. The number of males and females in the sample was essentially equivalent (Females = 29; 43.3%). Students were distributed across 1st through 6th grades, with most students in 2nd (n = 19; 28.4%) and 4th grades (n = 17; 25.4%). About half of the students were white (n = 33; 49.3%), with the remainder identifying as black (n = 27; 40.3%) and Hispanic (n = 5; 7.5%). The sample did not include any students identified as English Language Learners. Four students were receiving special education services for specific learning disability. The remaining students were identified as needing support through the following methods: teacher referral and less than 25th percentile on screening (n = 24, 35.8%), teacher referral only (*n* = 26, 38.8%), less than 30th percentile on screening (*n* = 6, 9.4%), teacher referral and less than 45th percentile on screening (*n* = 5, 7.9%), school psychologist referral (*n* = 4, 6.0%), and less than 25th percentile on screening (*n* = 2, 3.0%).

Setting Characteristics

Studies were conducted in the following areas of the United States: Northeast (*n* = 4; 25.0%), Southeast (*n* = 5; 31.3%), Midwest (*n* = 2; 12.5%), and South (*n* = 2; 12.5%). Two studies (12.5%) did not report the geographic location of the study and one study (6.3%) was conducted in Turkey. Seven studies (43.8%) did not report the urbanicity of the school. One-fourth of the studies were conducted in urban settings (*n* = 4), with the remainder in suburban (*n* = 2; 12.5%), rural (*n* = 2; 12.5%), and other settings (*n* =1; 6.3%). Studies reported the percent of students in the school that qualified for free/reduced lunch, which ranged from 42 to 90%; however, they did not report the free/reduced lunch status of the actual participants.

BEA Methodology

Table 2 provides an overview of study-level BEA methodology of the included studies. Over half of the BEAs were conducted to determine a skill versus performance deficit (*n* = 10, 62.5%). The remainder were conducted to determine the best skill intervention (*n* = 3, 18.8%) or the best performance intervention (*n* = 3, 18.8%). A majority of the studies (*n* = 13; 81.3%) utilized an abridged BEA as the primary methodology while 18.8% (*n* = 3) utilized a standard BEA comprised of three or four sessions for each treatment condition. A subset of the abridged studies (*n* = 3; 18.8%) used a one-trial BEA as their primary methodology (i.e., Duhon et al., 2004, Gilbertson, 2001; Gilbertson et al., 2008).

All studies used a multielement design to conduct the BEA and 75% (*n* = 12) used a multielement design to conduct the extended analysis. Two studies (12.5%) used a multiple baseline design in the extended analysis, and two studies (12.5%) did not include an extended analysis. To determine if an intervention was more effective than baseline, eight studies (50%) used visual analysis, four studies (25%) used a 20% criterion, and the remaining studies used a 25% criterion (*n* = 1, 6.3%), 50% criterion (*n* = 1, 6.3%), mean performance difference (*n* = 1, 6.3%), and an unspecified percentage increase (*n* = 1, 6.3%). Additionally, Mong and Mong (2012) calculated the percentage of non-overlapping data points (PND), by dividing the number of non-overlapping points in baseline with the number of points in each treatment condition, to supplement visual analysis during the extended analysis.

To determine the best intervention within the BEA, just over half of the studies used visual analysis (*n* = 10; 62.5%), while the remaining used visual analysis + 20% criterion (*n* = 2; 12.5%) or a mean performance difference (*n* = 1; 6.3%). This component was not applicable for the three studies that used a one-trial abridged design (i.e., Duhon et al., 2004; Gilbertson, 2001; Gilbertson et al., 2008).

Dependent and Independent Variables

Most studies (*n* = 14; 87.5%) used digits correct per minute as the primary dependent variable, while the remaining studies used total digits correct (*n* = 1; 6.3%) or total number of problems correct (*n* = 1; 6.3%). Seventeen different independent variables were applied across the identified studies. The most commonly used performance interventions were contingent reward (*n* = 11; 68.8%) and performance or corrective feedback (n = 4; 25%). The most commonly used skill-based interventions were cover-copy-compare (n = 6; 37.5%) and Math to Mastery (n = 4; 25%).

Primary Outcome Measures

Table 3 provides a summary of the primary outcome measures that were used in the studies. All researchers developed their own outcome measures to evaluate the outcomes of the BEA and the extended analysis. Five studies (31.3%) used measures that were created by randomizing problems, using either Excel or a web-based randomizer. Four studies (25%) developed probe sets that were assigned to treatment conditions, which is consistent with the recommendations for outcome measures by Daly et al. (1996, 1997). In these studies, problems were sequenced randomly across probes, and problems were arranged so that the same problems did not repeat consecutively (e.g., 3 + 4 would not be presented two times in a row; Poncy & Skinner, 2011). Twenty-five percent of studies (n = 4) used the Math Worksheet Generator from www.interventioncentral.org and one study (6.3%) used the Mathematics Worksheet Factory from www.mathaids.com. Additionally, no researchers considered SEM when determining if an intervention was more effective than baseline or when making comparisons.

Student-Level Outcomes

Table 4 provides a summary of student-level outcomes. For the students who participated in a one-trial BEA, the majority were hypothesized to have a skill deficit in math computational fluency (n = 11; 68.8%), with the remaining were hypothesized to have a combined skill and performance deficit (n = 4; 25.0%) or a performance deficit (n = 1; 6.3%). Contingent reward was the most frequently effective performance intervention (n = 11, 18.3%), while cover-copy-compare (n = 5; 8.3%) and Math to Mastery (n = 5; 8.3%) were the most frequently effective skill-based interventions.

Extended Analysis Results

Researchers included an extended analysis for most students (n = 60; 89.6%). Two studies did not conduct an extended analysis (Atbasi & Sanir, 2018; Carson & Eckert, 2003) and two students did not participate in an extended analysis because their performance had reached mastery criteria prior to implementing the extended analysis (Kleinert, 2017; Ota, 2008). For just over two-thirds of participants, the results of the extended analysis matched the results of the BEA (n = 43; 62.7%). For two students (4.3%) where the results of the extended analysis did not match the BEA, the most effective intervention in the extended analysis was the second-best intervention from the BEA (Clark, 2013; Kleinert, 2017). For seven students (11.1%), there was no differentiation displayed between interventions compared during the extended analysis (Clark, 2013; Gilbertson, 2001; Kleinert, 2017). For two students (4.3%), the worst or the second worst intervention in the BEA outperformed the most effective intervention from the BEA in the extended analysis (Kleinert, 2017; Reisener et al., 2016). Finally, for one student (2.1%), there was no differentiation in the BEA; however, when all interventions from the BEA were implemented during the extended analysis, one intervention (the combined skill and performance intervention) emerged as the most effective (Kleinert, 2017).

The best intervention from the BEA was implemented in the extended analysis for all students who participated in an extended analysis. For 11 students (18.3%), the researchers implemented the best intervention, to evaluate the effects of the intervention over time. For ten students (16.7%), the researchers compared the effects of the best intervention to a baseline or control condition (e.g., Codding et al., 2009; Everett et al., 2016; Reisener et al., 2016). A comparison intervention from the BEA (e.g., all other interventions, worst intervention, student choice) was compared to the best intervention from the BEA for 39 students (65%). For 19% of the students (n = 20), the extended analysis was a comparison of a skill intervention to a performance intervention (or a combined skill/performance intervention) (e.g., Duhon et al., 2004; Kleinert, 2017). Silva (2017) compared the effects of the most effective intervention in the BEA to a student-chosen intervention in the extended analysis. Hofstader-Duke (2015) implemented the intervention identified as most effective in the BEA across two different stimuli (fluent facts and non-fluent facts). Kleinert (2017) evaluated the effects of the best skill intervention from the BEA, the best performance intervention from the BEA, and a combination of the best skill and performance intervention from the BEA (e.g., explicit timing, escape, and explicit timing + escape) to baseline. Finally, Mong and Mong (2012) evaluated the effects of all interventions to baseline.

Quality Indicators

Table 5 presents the quality indicator ratings for each of the studies included. The highest quality ratings were found for the researchers' description and quantification of the dependent variable (i.e., digits correct per minute) and the description and manipulation of the independent variable(s). Additionally, researchers generally described the students with adequate detail, but did not describe the setting of the BEA and/or extended analysis. The lowest ratings were found for replication of effects across participants and for social validity. While all but one study included a minimum of three participants (e.g., Atbasi & Sanir, 2018), the results of seven studies (46.7%) included less than three replications of effects. Most studies did not document features of social validity of the intervention or BEA procedures, such as student or teacher acceptability, feasibility, effectiveness, or continued used. Additionally, all studies were implemented by researchers, rather than typical intervention agents (such as teachers, school-based interventionists, or school psychologists).

Discussion

The purpose of this systematic review was to locate all empirical studies that used a BEA to determine the most effective math intervention for students. Sixteen studies, that included 67 participants, were located. The studies represented most geographic regions of the US; however, the ethnicity of included participants was primarily limited to students who were white or black, with just five Hispanic students and no students identified as English Language Learners. Only four students were identified as eligible for special education for a specific learning disability.

Math BEA Design

Authors identified different purposes for conducting a BEA, including intervention comparisons or determining the function of math fluency deficits. Over half of the BEAs implemented both skill and performance interventions; the remaining studies implemented either skill interventions or performance interventions in isolation. One study, Everett et al. (2016), conducted a BEA to evaluate the individual components of a multi-component intervention program (Math to Mastery). All studies used a multi-element design to conduct the BEA. Eighty-one percent of the BEAs were conducted using an abridged design, while 18.8% used a standard design. Forty-six percent included a return to baseline condition followed by the reimplementation of the best intervention, to verify the effects of the best intervention within the BEA procedures. Three of the abridged design studies implemented a one-trial BEA, where a single reward or reward plus instruction trial was delivered to determine if the student's math difficulties were due to a skill or performance deficit.

Seventeen different interventions were implemented across the studies, thereby deviating from the approach used in reading BEA studies, in which a suite of consistent intervention options is employed (Burns & Wagner, 2008). The most frequently used skill-based interventions were cover-copy-compare and Math-to-Mastery. For the most part, performance-based intervention conditions were based on positive reinforcement (e.g., contingent reward and feedback) and rarely were interventions that addressed negative reinforcement (e.g., break from a math task) included. Two studies used a combined skill and performance intervention by implementing both contingent reward and instruction (Gilbertson et al., 2008) or break from a math task and instruction (Kleinert, 2017). It is concerning that negative reinforcement conditions were not more often considered within the context of a BEA, given the correlation between difficulties with math, math anxiety, and motivation and engagement with practice opportunities. Specifically, it has been shown that students who report higher levels of math anxiety have lower levels of math mastery as well as lower motivation to practice math (Ashcraft & Krause, 2007).

Measurement of the Dependent Variable

All studies targeted computation, with digits correct per minute serving as the primary dependent variable. Eighty-six percent of studies used a single-skill CBM-M to measure the dependent variable, while the remainder use a multi-skill probe. However, the computation skill that was targeted in the studies widely varied, ranging from single-digit to complex computation (e.g., 3-digit by 3-digit addition with regrouping). To measure the primary outcome variable, four studies (Codding et al., 2009; Hofstader-Duke & Daly, 2014; Kleinert, 2017; and Silva, 2017) developed probe sets and assigned problems to treatment conditions. The remaining studies developed their own measures by randomizing problems or using a Math Worksheet Generator from either interventioncentral.org or mathaids.com. These latter studies did not assign problems to treatment conditions. Two problems arise from the use of randomized probes that did not include exclusive problems assigned to conditions. The first is the lack of overlap between intervention and assessment materials makes it difficult to determine whether the measures were accurately capturing the effects of the treatment or repeated exposure to the same problems (Daly et al., 1996; Daly et al., 1997). Second, research has demonstrated that randomized probes, such as those generated from interventioncentral.org, have moderate test-retest and alternate-form reliability, which is below the threshold (0.80) required for progress monitoring (Strait et al., 2015), thereby increasing the measurement error of the scores.

None of the authors of math BEA studies considered the potential impact of measurement error on their decisions. This is potentially problematic, because Burns et al. (2017) found that in only one in four reading BEAs did an intervention produce an effect greater than the SEM of the outcome measure. That said, single-skill CBM-M is considered to be a well-established form of CBM, with alternate-form reliability coefficients above the threshold needed for progress monitoring (0.80) and relatively low variance across probes, given the skill required across probes is homogenous (Christ et al., 2008). Specifically, 80% of the variance associated with CBM-M is attributed to either individual differences or developmental level, with very little of the remaining variance associated with differences between probes or unsystematic error (Hintze et al., 2002). However, it is possible that because probes with unknown variability (i.e., those constructed with random problem generators) were used across multiple studies, BEA decisions made based on single data points may reflect unsystematic error, rather than true intervention effects.

Different decision rules were applied across this set of studies, including a 20% criterion, 25% criterion, 50% criterion, mean performance difference, visual analysis, and visual analysis+20% criterion. The most commonly used decision rules were visual analysis and the 20% criterion. These results are consistent with research in the reading BEA literature, which has also used a 20% criterion to determine if an intervention is more effective than baseline and visual analysis to determine the most effective intervention overall (Daly et al., 1999; Jones & Wickstrom, 2002; Noell et al., 2001). It has been suggested that more liberal decision-making criteria, such as the 50% criterion may be problematic in that it may result in students being overidentified as having a performance-only deficit (Solomon et al., 2018). Recently, Soloman et al. (2018) conducted an empirical analysis to determine which decision-making criterion would be most appropriate within a skill versus performance analysis and found that the 20% criterion is generally the most appropriate. The findings of the current synthesis, along with the recent results of Soloman et al. (2018), suggest that in future math BEA studies, a 20% criterion may be the most appropriate and can be used to develop consistency in decision making procedures across math BEA studies.

BEA and Extended Analysis Outcomes

The most frequently effective skill-based interventions were cover-copy-compare and Math-to-Mastery and the most frequently effective performance-based intervention was contingent reward. Eighty-nine percent of studies conducted an extended analysis. Of these studies, 75% of studies used a multi-element design and two studies used a multiple baseline design during the extended analysis. For 71.7% of students who participated in an extended analysis, the results matched the results of the BEA. It cannot be determined if a standard or abridged BEA was better in predicting the outcomes of an extended analysis as only one of the two studies that conducted a standard BEA included an extended analysis. When the extended analysis results did not match the intervention identified in the BEA, there was either a lack of differentiation in the extended analysis (58.3%), the second best intervention in the BEA was the most effective (16.7%), the worst intervention in the BEA was the most effective (16.7%), or there was a lack of differentiation in both the BEA and in the extended analysis (8.3%). When there was a lack of differentiation, authors concluded that the different treatment conditions produced similar responding (Kleinert, 2017) or chose the intervention with the greatest overall mean improvement (Clark, 2013). It is possible that a mismatch between the BEA and the extended analysis could have been a result of a decision error made in BEA, due to measurement error and/or a lack of overlap between instructional and assessment materials.

Previous research has indicated that the extended analysis can serve multiple purposes, including extending the BEA when results are inconclusive (McComas et al., 1996; Vollmer et al., 1995), examining the results of the best intervention in the BEA compared to baseline over time (Jones & Wickstrom, 2002), and determining the generalizability of the best intervention (Daly et al., 2006). Overall, the results of the current synthesis indicate that researchers implemented a variety of treatment combinations during the extended analysis, with the most common combinations being a comparison of a skill-based intervention to a performance-based intervention after a one-trial BEA (20%), implementing the best intervention only over time (18.3%), the best intervention from the BEA to a control condition (16.7%), all interventions from the BEA (13.3%). Recent research has called for more rigorous methods of testing intervention effectiveness than simply comparing an intervention to a baseline or business-as-usual condition (Kilgus et al., 2016). Specifically, researchers have been called on to assess the efficacy of interventions by comparing the effects of an intervention to another intervention or a gold standard practice, to establish a higher standard for intervention performance (Kilgus et al., 2016).

Study Quality

Quality indicators, which were based on the single-case design criteria outlined by Horner et al. (2005) and adapted to include recommendations for alternating treatment designs by *What Works Clearinghouse*, were applied to this study. Areas for improvement include measuring aspects of social validity and replication of intervention effects. Specifically, single case design researchers need to conduct direct replication, systematic replication and clinical replication of original study methods and procedures

(Kratochwill et al., 2018). Replication of single case design studies, with the goal of replicating both positive and negative results, is important for determining the efficacy of an intervention, determining if an intervention is effective over time, and enhancing the credibility of the intervention when replicated across different samples, sites, and research groups (Kratochwill et al., 2018). With respect to social validity, researchers should consider the feasibility of BEA as an assessment tool as well as student or teacher acceptability of the most effective intervention in the BEA. This is important, because if the social validity of either the BEA or the chosen intervention is low, it is unlikely that either will be implemented in the future (Kratochwill et al., 2018). Finally, most of the studies met full criteria for the dependent variable as outlined by Horner et al. (2005) (i.e., the dependent variable is operationally defined and measured repeatedly, inter-observer agreement is calculated, and the social significance of the dependent variable is considered). However, it should be noted that only four studies (25%) developed probe sets that were assigned to treatment conditions, which is consistent with the recommendations for measurement of the dependent variable in a BEA by Daly et al. (1996, 1997).

Future Research and Practice Implications

The results of this synthesis provide a starting point for better understanding the use of BEA to determine an effective math intervention. The results also provide direction for future research and practical application. First, it should be noted that the research on BEAs in math has been conducted with a small number of students with disabilities (6%) and with a narrow subset of the population (i.e., white and black students), limiting the extent to which these results can be generalized to a more diverse student population. Second, further work needs to be done on the development of probe sets that are accessible to researchers and practitioners and can be used to evaluate intervention effectiveness across multi-element designs. Without established, reliable, and available probe sets it is unlikely that practicing school professionals will employ a BEA in math. That is, it is much easier to locate free online worksheet generators for various skills and use these tools as the primary outcome variable to test different intervention conditions. However, while widely used (e.g., the math worksheet generator on interventioncentral.org was visited 121,562 times between 2014 and 2015; Strait et al., 2015), educators know very little about the technical adequacy of the probes that are generated and used (Strait et al., 2015). Furthermore, replicating the use of a consistent decision-making criteria (i.e., the 20% criterion) to determine an effective intervention compared to baseline is warranted. Through the use of meta-analytic procedures, researchers demonstrated that in the reading BEA literature, an average no-assumptions effect size of 2.80 and a percent of non-overlapping data (PND) of 80% between the interventions tested within the BEA, may be used to determine the most effective reading fluency intervention (Burns & Wagner, 2008). These criteria provide direction for future research on decision making in math BEAs.

Finally, replication of math BEAs that use a standard suite of interventions, consisting of interventions, which may include contingent reward, cover-copy-compare, and Math-to-Mastery, is warranted. Doing so will further validate this suite as a way of distinguishing between skill and performance-based deficits in math computational fluency, allow for replication of the same suite of interventions across multiple participants, increase the generalizability of math BEA results, and allow for the opportunity to further synthesize the math BEA literature. Finally, to this date, math BEA studies have only been implemented by researchers. Studies should begin to examine the social validity of the use of BEAs by including natural implementers, such as practicing school psychologists or special educators.

Conclusion

Sixteen studies with 67 students have applied a math BEA to improve computation skills, across a wide variety of whole number operations. Eighty-six percent of the studies used an abridged-BEA design to determine the best intervention, and for two-thirds of the cases, an extended analysis verified the results. There is considerable interest in bolstering school psychologists' skills in using data to make instructional decisions for students (Burns & Coolong-Chaffin, 2006; Codding et al., 2009; Martens & Gertz, 2009) and BEA has the promise of being a method of doing just that (see Lemons et al., 2018). However, it is currently difficult to make recommendations for the use of BEAs for math in practice, without further research to validate the technical properties of this method. With additional systematic research, we can begin to recommend BEA as a valid method for determining instructional needs for students and make recommendations for the use of this method in practice.

Table 1

Participant and Setting Characteristics of Included Studies

Variable	*n*	%
Total Number of Participants	67	
Female	29	43.3
Male	38	56.7
Grade		
1st	6	9.0
2nd	19	28.4
3rd	9	13.4
4th	17	25.4
5th	5	7.5
6th	8	11.9
Not Reported	3	4.5
Race		
White	33	49.3
Black	27	40.3
Hispanic	5	7.5
Not Reported	2	3.0
English Language Learner	0	0.0
Eligible for Special Education	4	6.0
Geographical Location		
Northeast	4	25.0
Southeast	5	31.3
Midwest	2	12.5
South	2	12.5
Turkey	1	6.3
Not Reported	2	12.5
Urbanicity		
Not Reported	7	43.8
Urban	4	25.0
Suburban	2	12.5
Rural	2	12.5
Other	1	6.3
Percent Free/Reduced Lunch of School		42 - 90

Note. Reisener et al (2016) included two studies in a single manuscript. Each study was counted separately, resulting in a total of 16 studies, with 67 total participants.

Table 2

Study Level BEA Methodology

Variable	*n*	%
Purpose for BEA		
Skill vs. Performance	10	62.5
Best Skill Intervention	3	18.8
Best Performance Intervention	3	18.8
Type of BEA		
Abridged	13	81.3
Standard	3	18.8
Experimental Design of BEA		
Multielement	16	100.0
Experimental Design of Extended Analysis		
Multielement	12	75.0
Multiple Baseline	2	12.5
Not Applicable	2	12.5
Baseline Criteria Used		
Visual Analysis	8	50.0
20% Criterion	4	25.0
25% Criterion	1	6.3
50% Criterion	1	6.3
Mean Performance	1	6.3
Percent Increase	1	6.3
Best Intervention Criteria Used		
Visual Analysis	10	62.5
Visual Analysis + 20% Criterion	2	12.5
Mean Performance	1	6.3
Not Applicable	3	18.8
Primary Dependent Variable		
Digits Correct per Minute	14	87.5
Total Correct Digits	1	6.3
Total Problems Correct	1	6.3
Independent Variable(s)		
Contingent Reward	11	68.8
Cover-Copy-Compare	6	37.5
Math to Mastery	4	25.0
Feedback (Performance/Corrective)	4	25.0

Constant Time Delay	3	18.8
Student Choice	3	18.8
Timed Sprint	3	18.8
Contingent Reward + Instruction	2	12.5
Goal Setting	2	12.5
Taped Problems	2	12.5
Repeated Practice	2	12.5
Repeated Practice + Contingent Reward	1	6.3
Cover Copy Compare + Repeated Practice	1	6.3
Escape	1	6.3
Explicit Timing	1	6.3
Attention (Experimenter/Peer)	1	6.3
Folding-In Technique	1	6.3

Note. Total number of studies included is 16. IV percentages are greater than 100%, given more than one IV was included per study. The *n* for Extended Analysis Comparison includes only studies that conducted an extended analysis, which is 14.

Table 3

Primary Outcome Measure

Variable	*n*	%
Type of Measure Used		
Researcher Developed	16	100.0
Standardized	0	0.0
Type of Researcher Developed Measure		
Randomized problems	5	31.3
Probe sets	4	25.0
Intervention Central Worksheet	4	25.0
Math Aids Worksheet	2	12.5
Mathematics Worksheet Factory	1	6.3
Problems Assigned to Treatment Conditions		
Yes	5	31.3
No	11	68.8
Researchers Ensured Probe Equivalency		
Yes	4	25.0
No	12	75.0
Consideration of SEM		
Yes	0	0.0
No	16	100.0

Note. One research team did not assign problems to treatment conditions; however, they did assign problems to the worksheets used.

Table 4

Student-Level BEA Outcomes

Variable	*n*	%
One Trial BEA Hypothesized Deficit		
Skill Deficit	11	68.8
Skill and Performance Deficit	4	25.0
Performance Deficit	1	6.3
Best Skill Intervention		
Reward	11	18.3
Cover-Copy-Compare	5	8.3.
Math to Mastery – All components	5	8.3.
Contingent Reward + Instruction	4	6.7
Taped Problems	4	6.7
Timed Sprint	4	6.7
Explicit Timing + Escape	3	5.0
Repeated Practice	2	3.3
Experimenter Attention	2	3.3
Constant Time Delay	2	3.3
Corrective Feedback	2	3.3
Folding-In Technique	2	3.3
Goal Setting	2	3.3
Math to Mastery – 3 components	2	3.3
Repeated Practice	2	3.3
Escape	1	1.7
Explicit Timing + Reward	1	1.7
Student Choice + Verbal Encouragement	1	1.7
Extended Analysis Comparisons		
Skill vs Performance	12	20.0
Best intervention only	11	18.3
Best vs Baseline/Control	10	16.7
All interventions	8	13.3
Best vs Worst	6	10.0
Best vs Student Choice vs Control	5	8.3
Skill vs Performance vs Skill + Performance vs Baseline	5	8.3
All interventions and Baseline/Control	3	5.0
Not included	7	11.7

Extended Analysis Included		
Yes	60	89.6
No	7	11.1
Extended Analysis Matched BEA		
Yes	43	71.7
No	12	20.0
IV of Extended Analysis Not Defined	5	8.3

Note. 16 of the 67 included students participated in a Can't Do/Won't Do assessment, while the remainder participated in complete BEAs. Therefore, the total *n* for the Can't Do/Won't Do results is 16 and the total *n* for the remainder of the BEA variables is 51.

Table 5

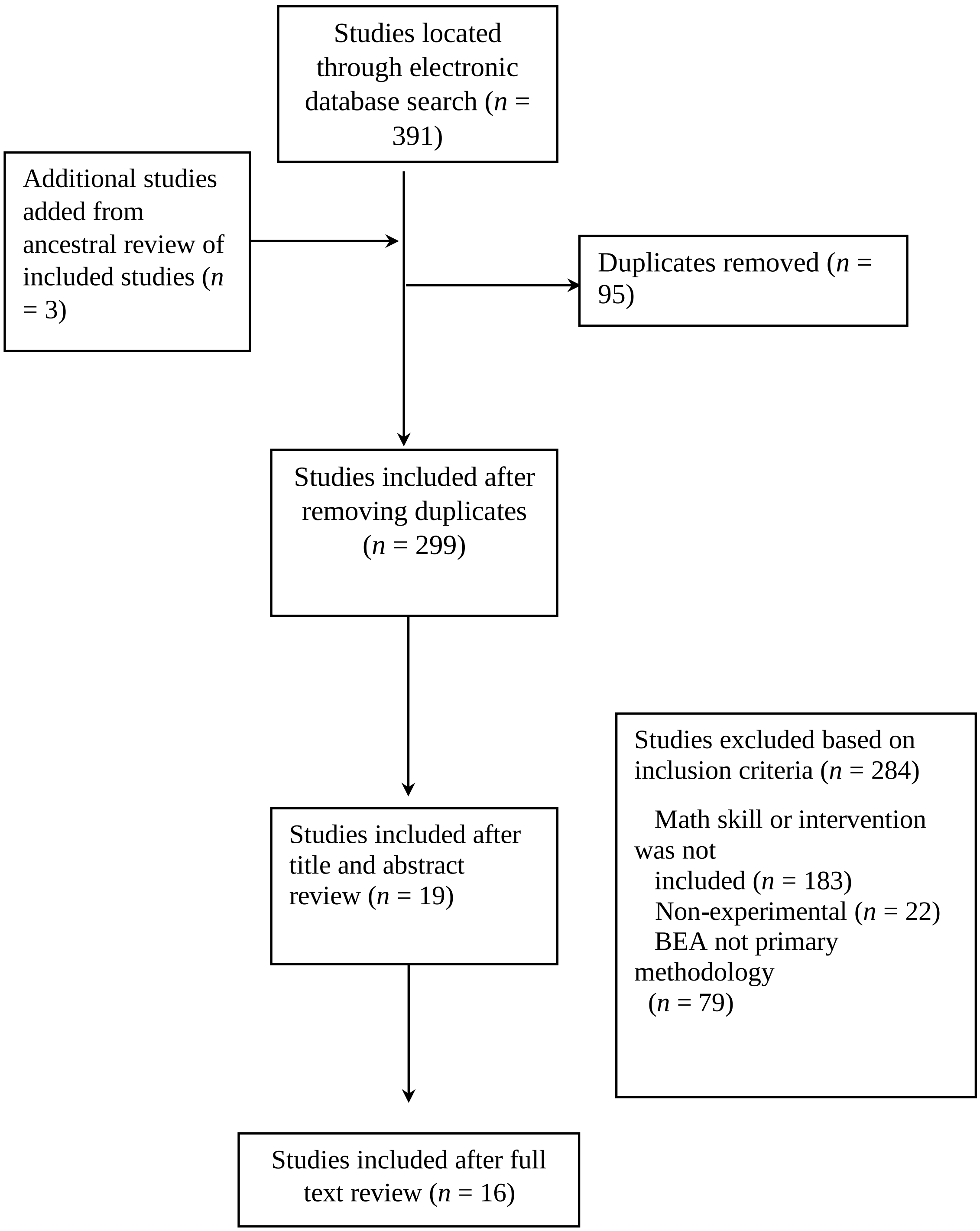
Mean Quality Indicators of Included Studies

Study	Participants and Setting	DV	IV	Baseline	Extended Analysis	Internal Validity	Replication	Social Validity
Atbasi & Sanir (2018)	3	3	3	3	NA	2.67	2	2
Carson & Eckert (2003)	3	3	3	3	NA	3	3	2
Clark (2013)	2.67	3	3	3	2.5	2	2	2
Codding et al. (2009)	3	3	3	3	3	2.67	3	2
Duhon et al. (2004)	2.67	3	2.33	1	3	3	NA	2.25
Everett et al. (2016)	2.33	2.5	3	3	3	3	2	2.5
Gilbertson (2001)	2	2.25	2.33	3	2	2.5	3	2
Gilbertson et al. (2008)	2.33	2.75	3	3	3	3	3	2.75
Hofstadter-Duke & Daly (2015)	2.67	3	3	1	2.5	2	3	2.25
Kleinert (2017)	2.67	3	3	3	2.5	2.67	3	2.25
Ota (2008)	3	3	3	3	1.5	2	2	2.5
Mellot & Ardoin (2019)	2	2.75	3	3	2.5	2.33	3	2
Mong & Mong (2012)	3	3	3	3	2.5	3	3	2.5
Reisener et al. (2016) Study 1	3	3	3	3	3	2.33	2	2.25
Reisener et al. (2016) Study 2	3	3	3	3	2.5	2.33	2	2.25
Silva (2017)	3	3	3	3	2.5	2	2	2.75
MEAN	2.71	2.9	2.9	2.75	2.57	2.47	2.53	2.27

Note. Ratings are based on an adaptation of (Xin, 2008). 1 = indicator not met, 2 = indicator partially met, and 3 = indicator met.

Figure 1

Process of Article Inclusion and Exclusion



Chapter 3

A Gated Screening Approach for Instructional Planning in Whole Number Proficiency

Universal screening sets the foundation for a successful Response to Intervention (RtI) system by identifying students who are at risk for reading and math (Fuchs & Vaughn, 2012). When conducting universal screening in reading, schools can administer a measure of oral reading fluency (CBM-R). CBM-R is a robust, brief, general outcome measure of overall reading competence (Fuchs et al., 2001; Hasbrouck & Tindal, 2006; Shinn, 1989), is predictive of later achievement (Hintze & Silberglitt, 2005), and can be used across different grades to reflect improved literacy skills (VanDerHeyden et al., 2019) as well as for instructional planning (Szadokierski et al., 2017). Conducting universal screening in math is more complicated because math proficiency cannot be measured in a single, robust general outcome measure (similar to CBM-R; Foegen et al., 2007). Rather, math proficiency is skill specific and reflects mastery of grade-level content (Stecker et al., 2005). Math content is multifaceted and requires the acquisition of increasingly difficult content, making it impossible to assess using a single skill over time (e.g., Foegen et al., 2007). Therefore, different measures perform better at different grade levels for identifying who is at risk for not meeting end-of-year standards (VanDerHeyden et al., 2017). Furthermore, few studies have examined how to use universal screening measures for the purpose of instructional planning, which is necessary to provide a seamless transition from screening to intervention for those who are identified as at risk. Therefore, the purpose of this study was to evaluate a framework for math screening that could identify students as at risk and aid in instructional planning.

Current Screening Recommendations in Math

There are several recommendations for schools to consider when implementing universal screening in math. The Institute on Education Sciences (IES) recommends that schools (a) set up a screening system with measures that are efficient, reasonably reliable, demonstrate predictive validity, and are conducted in the beginning and middle of the school year; (b) choose measures that cover the instructional objectives for each grade; (c) use screening data along with state testing results in Grades 4 through 8; and (d) use the same system in all schools, so that results can be aggregated across schools and analyzed at the district level (Gersten et al., 2011).

Many studies have examined which math measures are the most valid and reliable to use for universal screening and have supported the recommendation that when previous state test scores are available (e.g., Grade 4 and above), the preceding year's test score is a viable measure of determining risk status (Klingbeil et al., 2019; Nelson et al., 2016, 2017; VanDerHeyden et al., 2017). In Grades 4 and 5, the combination of the preceding year's test score with AIMSweb™ Math Concepts and Applications (M-CAP) offers the strongest prediction, while in Grade 3, where state testing data is not available, AIMSweb™ Math Computation (MCOMP) offers the strongest prediction (VanDerHeyden et al., 2017). However, AIMSweb™ MCOMP yields a high false-negative error rate, meaning it may fail to accurately identify all students who are actually at-risk (VanDerHeyden et al., 2017). An alternative option is the use of computer-adaptive tests, such as STAR Math (Renaissance Learning, 2016), which may have higher levels of sensitivity and specificity than multiple-skill curriculum-based measures in predicting students' math achievement levels (Shapiro & Gebhart, 2012). Additionally, a single-skill computation probe, referred to as a subskill mastery measure, that measures key grade-level computation skills may also be an acceptable alternative in some grades (e.g., Nelson et al., 2017; VanDerHeyden et al., 2017).

These recommendations are useful for identifying students who are at-risk for not meeting end-of-year grade-level standards, but they do not lend themselves to informing intervention to students who may need supplemental instruction. Additionally, the recommended measures evaluate the broader math content, which does lend them to have stronger prediction of future achievement (e.g., AIMSweb™ M-CAP), but these broader measures do not provide sufficient information to effectively identify what specific math skills a student is struggling in (Foegen, 2007). Rather, these broader measures can be used for instructional grouping and to indicate a starting point for more detailed, diagnostic assessment to inform instructional planning (Hosp & Ardoin, 2008).

Many educators express frustration with the need to conduct a diagnostic assessment prior to assigning a student to a supplemental intervention, because it can be time and resource intensive (Hosp & Ardoin, 2008). However, if schools do not take the time to conduct a more detailed, diagnostic assessment to make instructional decisions, students who need supplemental intervention may be engaged in valuable instructional time that does not actually address their needs (Hosp & Ardoin, 2008). In practice, this has resulted in RtI being referred to as a “wait-to-fail” model, where students may receive an intervention that was not likely to be effective for them, based on their prerequisite skills (Fuchs & Vaughn, 2012). Therefore, it is important to consider ways to make math universal screening not only relevant for determining risk, but for also for providing data that can aid in effective instructional planning. Doing so would eliminate the need for a more time-consuming diagnostic assessment process and the wasting of precious resources, such as time being used on ineffective interventions (Nelson et al., 2017).

Gated Screening

One way of making universal screening relevant for accurately determining risk and for instructional planning is to conduct gated screening. Gated screening is a multivariate universal screening framework that uses two or more measures to improve the identification of students who are at risk or struggling to meet end-of-year grade-level standards (Compton et al., 2010). Recently, there appears to be consensus around the need to agree on recommendations for gated screening in math. These recommendations are intended to verify a student's at-risk status, so that only students who are actually at-risk receive supplemental intervention and include using previous end-of-year test scores as the first gate and a universal screening measure as the second gate (Nelson et al., 2016; Van Norman et al., 2017; Van Norman et al., 2018; VanDerHeyden et al., 2017). However, Van Norman et al. (2018) indicated that a gated screening process could also be used for instructional planning in math, by using a strong initial screening measure with a high cut score (e.g., 40th or 50th percentile) in the first gate and a criterion-referenced or subskill mastery measure, such as a measure of whole number proficiency, as the second gate (Vaughn & Fletcher, 2012). Including a subskill mastery measure in the second gate can verify that a student is at-risk and provide useful data for instructional planning (Van Norman et al., 2018).

Subskill mastery measures can provide more detailed, diagnostic information to identify a student's specific skill strengths and weaknesses (Hosp & Ardoin, 2008; Shinn & Bamonto, 1998). Subskill mastery measures can also inform whether a student will require an acquisition- or fluency-building intervention (Shapiro, 2011). Therefore, following the recommendation to include a subskill mastery measure in the second gate will likely lead to information that can be used to match a student to an intervention that meets their needs (Van Norman et al., 2018). This will provide a seamless transition from screening to intervention, meaning that time is not lost between when a student is identified as at risk and when an effective intervention is implemented.

Whole Number Proficiency

In elementary grades, it may be important to include a subskill mastery measure of whole number proficiency in the second gate, given whole number proficiency is an indicator for the development of more complex math skills (e.g., solving complex problems and interpreting abstract mathematical concepts; Patton et al., 2017) and also a predictor of outcomes on state assessments (Shapiro et al., 2006). Whole number proficiency, defined as the efficient and accurate completion of math calculation (National Council of Teachers of Math [NCTM], 2008) is one of the important curricular content areas recommended by the National Mathematics Advisory Panel (NMAP, 2008). Whole number proficiency is a complex skill that depends on a student's attentive behavior, reasoning skills, central executive skills (e.g., working memory), and early numeracy skills (e.g., flexible use of counting procedures; Fuchs et al., 2016), in addition to early literacy skills (e.g., phonological processing; Fuchs et al., 2019). Whole number proficiency also requires a student to be sufficiently accurate as well as fluent, both of which are substantial predictors of later word-problem solving and pre-algebraic knowledge (Fuchs et al., 2016). Math achievement trajectories are established early (i.e., before Grade 4), so it is important for educators to assess whole number proficiency before fourth grade and to provide effective intervention to prevent difficulty with more complex mathematical skills (Fuchs et al., 2016).

The Common Core State Standards (2010) also recommend that students demonstrate accurate addition and subtraction skills within 20 by the end of Grade 2 and the National Mathematics Advisory Panel recommends that students demonstrate fluent retrieval of addition and subtraction facts by the end of Grade 3. However, research has demonstrated that only 50% of Grade 3 students were able to demonstrate fluent retrieval of addition facts while only 26% were able to demonstrate fluent retrieval of subtraction facts (Stickney et al., 2012). These findings further support the need to provide appropriately matched, early intervention in the area of whole number proficiency.

Instructional Planning in Math

The use of a subskill mastery measure of whole number proficiency in a gated screening framework is supported by theories of instructional planning. Using assessment data for instructional planning involves not only the determination of what-to-teach, but also how-to-teach (Hosp & Ardoin, 2008; Zigmond & Miller, 1986). Determining what to teach includes identifying the skills a student has mastered and the skills a student has yet to master (Hosp & Ardoin, 2008). The skills that are assessed need to be teachable and serve as a prerequisite skill for a more complex skill (e.g., whole number proficiency; Hosp & Ardoin, 2008). This information can be gathered through the use of subskill mastery measures, which represents a breakdown of curriculum outcomes by subskills (Fuchs & Deno, 1991).

Instructional planning also involves determining how-to-teach, which refers to the types of instructional procedures that are needed to improve students' skills (Hosp &

Ardoin, 2008). The instructional hierarchy (Haring & Eaton, 1978) can be used to theoretically understand how a students' skills are progressing through four main stages: acquisition, proficiency, generalization, and adaption. The instructional hierarchy also provides instructional recommendations based on the stage of learning that the student is in (Haring & Eaton, 1978). For example, students who exhibit whole number proficiency skills within the acquisition range are likely to benefit from instruction that includes modeling, guided practice, immediate feedback, and error correction procedures (Ardoin & Daly, 2007; Lovitt, 1978).

Research has shown that in math, the transition between the stages of acquisition and fluency are indicated by a fluency metric, or the number of digits correct per minute (VanDerHeyden & Burns, 2018). Specifically, in Grades 2 and 3, students who can complete below 14 digits correct per minute are hypothesized to be within an acquisition level of learning, while students who can complete above 14 digits correct per minute are hypothesized to be within a fluency level of learning (Burns et al., 2006). Therefore, it is likely that different instructional procedures would be effective, depending on how fluent a student is prior to receiving intervention. Two studies in math have demonstrated this, indicating that initial score on a subskill mastery measure can predict whether an acquisition or fluency based intervention is more suitable. Codding et al. (2007) found that students who are within the acquisition stage of learning typically best respond to modeling, immediate feedback, and error correction strategies, while students who are within the fluency stage of learning typically benefit most from repeated and timed practice. Additionally, a meta-analysis on the effectiveness of acquisition and fluency interventions found that acquisition interventions resulted in larger effect sizes among children with acquisition level skills, but only moderate effects for students with fluency level skills (Burns et al., 2010).

Determining how-to-teach also involves conducting trial teaching sessions, to validate the hypotheses that are made about the types of instructional procedures that will be most effective. Trial teaching sessions have been conceptualized as brief experimental analyses (BEA; Daly et al., 1996; 1997; Ysseldyke & Alogozzine, 1984). One version of using a BEA for trial teaching includes a can't do/won't do (CDWD) assessment, which can be used to determine if a student's academic difficulties are due to a skill (can't do) or performance (won't do) deficit (Codding et al., 2009; McKevett & Codding, 2019; VanDerHeyden & Witt, 2007). Ardoin et al. (2005) used a CDWD assessment within a screening model in math to verify if students who were identified as at-risk were an appropriate match for the skill-based interventions that were available in the school. There are also expanded versions of a brief experimental analysis, which include comparisons of different intervention conditions (e.g., Mong & Mong, 2012) or intervention tactics that range in intensity (e.g., Everett et al., 2016). Within the context of RtI, BEAs can be used to empirically validate the hypotheses made based on a student's preintervention performance, ensuring that intervention resources are appropriately allocated to student needs (VanDerHeyden & Burns, 2009).

Despite the useful instructional planning information that can be derived from a BEA, the procedure is time and resource intensive as it can take as much as 90-min to complete with each student (Jones & Wickstrom, 2002). Unfortunately, many educators who are implementing RTI indicate that they have needed professional development and training on how to use data to guide decisions for instruction and intervention.

Insufficient teacher training and a lack of resources, data and knowledge regarding intervention and instruction, progress monitoring and data use are also the top obstacles to adequate RTI implementation (Spectrum K-12, 2010). Combined with survey data reporting that intervention planning and preparation activities are not typically part of regular practice in schools (Silva et al., 2020), there is a need to investigate feasible and usable alternative methods for instructional planning in order for students to access effective interventions quicker.

Using a sample of 49 second- and third-grade students, Szadokierski et al. (2017) evaluated whether CBM-R could be used to predict effective intervention. Based on the student's performance on CBM-R and using the instructional hierarchy as a theoretical framework, Szadokierski and colleagues' hypothesized whether an acquisition-based intervention (modeling with error correction) or a fluency-based intervention (repeated reading with reward) would be most effective (Szadokierski et al., 2017). This study used a BEA to conduct trial teaching sessions and to validate intervention predications. Results indicated that there was a statistically significant difference in performance on CBM-R between students for whom modeling with error correction and repeated reading with reward was most effective, and performance on CBM-R accurately predicted which intervention would be most effective (Szadokierski et al., 2017). This study provides a model for a simplified screening framework that can be used to simultaneously identify students who would benefit from additional supports and accurately assign at-risk students to a supplemental intervention. The current study will evaluate similar methods to identify students who are at-risk in math and assign them to an effective intervention;

however, because of the unique nature of conducting universal screening in math, a gated screening framework will be used.

Purpose

Conducting universal screening is a necessary and commonly occurring component of effective implementation of RtI in math (VanDerHeyden et al., 2013). However, there is a need for additional guidance on how to use universal screening data in math for instructional planning, which would allow educators to go from screening to intervention in an accurate and efficient manner. The primary purpose of this study was to evaluate a gated screening framework in math that could be used for instructional planning for students who are at-risk in whole number proficiency. STAR Math and AIMSweb^{TM} MCOMP, two universal screening measures, were used to identify students who were at risk. Similar to the methods of Ardoin et al. (2005), a can't do/won't do assessment, using the AIMSweb^{TM} Subskill Mastery Measure Add/Sub (SSMM-Add/Sub), was used to verify whether students identified as needing a supplemental intervention in whole number proficiency were an appropriate match for the skill-based interventions that were available. As part of instructional planning, a BEA was implemented to verify which intervention was most effective for each student (modeling with error correction or explicit timing with reward). Then, it was determined whether each of the included screening measures could be used to accurately differentiate between the students who benefitted the most from each intervention and accurately predict the outcomes of the BEA. The research questions guiding this study were:

1. Are there statistically significant differences in students' performance on the measures included in the gated screening framework (STAR Math, MCOMP,

SSMM-Add/Sub), between those for whom modeling with error correction was most effective and those for whom explicit timing with reward was most effective?

2. Which variables in the gated screening framework (STAR Math, MCOMP, can't do/won't do assessment, and SSMM-Add/Sub) best predicted which intervention was the most effective in the BEA?

The current study also addressed an exploratory research question:

3. What are the cut scores on the variables that are significantly different between each effective intervention group, to provide enough specificity and sensitivity to predict which intervention a student will respond best to?

Method

Participants and Setting

The results of an a priori power analysis, conducted using G-power, indicated that 65 to 87 participants would be sufficient for detecting a moderate to large effect at a .05 significance level. Burns et al. (2010) found that acquisition mathematic interventions produce large effects with children whose skills are at a frustrational level; however, they produce only moderate effects for children whose skills are within the instructional range (Burns et al., 2010). Because there was a small number of studies that included students with instructional-level skills, Burns et al. (2010) could not make definitive conclusions about the effect size. Therefore, a medium to large effect size was determined to be most appropriate for this study.

One rural school district located in the Upper Midwest was selected using convenience sampling. Two hundred and thirty-seven students in the school (127 second grade; 110 third grade), were screened for eligibility to participate using grade-level AIMSweb^{TM} Math Computation (MCOMP) probes (Pearson, 2012). Screening was conducted at the classroom level on two consecutive days (one day for third grade and the next day for second grade). Students who scored at the 45th percentile were eligible to participate in the next phase of the study, as this was the percentile cutoff set by Pearson (the publisher of the AIMSweb^{TM} screening measure) to identify students who would benefit from Tier 2 supports. This cut score was also hypothesized to reduce the false negative rate, ensuring that all students who needed supplemental intervention were identified (Van Norman et al., 2018). Eighty-three students (35% of screened students) were eligible to participate (45 second grade; 38 third grade). Active consent forms were sent home to the parents/guardians of all eligible students and eight students (two second grade; six third grade) did not receive consent to participate. Student assent was obtained from all students who received parental consent. In total, 75 students participated.

Study participants were 43 (58.4%) second- and 32 (41.6%) third-grade students. Forty-two participants identified as White (57%), 29 identified as Native American (39%), one identified as Black (1.3%), and two identified as Hispanic (2.6%). Thirteen participants (16.9%) received special education services. One participant was eligible under hearing (1.3%), one was eligible under specific learning disability in basic reading skills (1.3%), and one was eligible under other health impairment (1.3%). Two participants were eligible under intellectual disability (2.6%), three were eligible under significant developmental delay (3.9%), and five were eligible under speech and language (6.5%). Additionally, 70.1% of participants were eligible for free or reduced priced lunch. The primary language of all students was English.

Dependent Measures

Three measures were used in the gated screening framework, AIMSweb™ MCOMP, Star Math, and AIMSweb™ SSMM-Add/Sub. Probes were also created for progress monitoring in the BEA. Social validity measures, the Children's Intervention Rating Profile (CIRP; Witt & Elliott, 1985) and an adapted version of the Intervention Rating Profile-15 (IRP-15; Witt & Martens, 1983), were used to gather information on students' and teachers' acceptability of each intervention that was implemented in the BEA.

Screening Measures

Grade-level AIMSweb™ MCOMP was used as the first screening measure and to determine which students would be eligible to participate in the study (Pearson, 2012). The participating school also used STAR Math (Renaissance Learning, 2016) as a universal screening measure, so this data was collected and analyzed as part of the gated screening framework. AIMSweb™ SSMM-Add/Sub was used for the can't do/won't do assessment (VanDerHeyden & Witt, 2007), to determine if the eligible students would be appropriate for the skill-based interventions that were available. Basic whole number addition and subtraction computation skills were targeted in this study, given this skill is foundational for the mastery of more difficult skills (Fuchs et al., 2016) and the ability to add and subtract within 20 is a Common Core State Standard in Grade 2 (National Governors Association, 2010).

MCOMP. MCOMP is a measure of grade-level computation skills, developed by AIMSweb™. The second-grade version of MCOMP assesses single-digit and double-digit addition and subtraction (with and without regrouping) and addition of three single-digit numbers. The third-grade version of MCOMP assesses single-digit, double-digit, and triple-digit addition and subtraction (with and without regrouping), addition of three single-digit numbers, single-digit multiplication, and division by a single digit. Students are given 8-min to complete as many problems as they can. MCOMP is scored using an answer key provided by AIMSweb™ that shows the correct answer to each problem and the number of points for each correct response; some problems are worth more points than others. The MCOMP total score was used for data analysis. MCOMP reliability estimates are .82 (SE = 4.8) for second grade and .89 (SE = 5.8 for third grade. MCOMP criterion validity is .84 for first grade and .73 for third grade.

STAR Math. For students who were eligible to participate in the study based on their performance on the MCOMP, the school also provided winter STAR Math scores to include as part of data analysis. STAR Math is a computer adaptive test developed by Renaissance Learning (2016), that measures skill development in number and operations, algebra, geometry, measurement, data analysis, statistics and probability. Test items are aligned to national and state standards, item exposure is controlled based on student responses, and students respond to a total of 34 questions per session. Test-retest and alternate form reliabilities range from .77 to .82 and the reliability of growth ranges from .71 to .74. Correlations with end-of-year state tests range from .63 to .80 (Renaissance Learning, 2016).

SSMM-Add/Sub. The SSMM-Add/Sub curriculum-based measure was used as part of the can’t do/won’t do assessment. The SSMM-Add/Sub includes 84 addition and subtraction 0 to 12 fact families. Problems are arranged across two pages in six rows and seven columns. Students are given 2-min to solve as many addition and subtraction problems as they can. The number of correct digits that a student wrote in the answer was scored and summed. A correct digit was counted as a correct number written in the correct place value. An incorrect digit was counted as an incorrect number written in the place value (Shinn, 2004). If a reversed number was obvious, but correct, it was counted as a correct digit. However, if a digit was reversed or rotated and created an incorrect number (e.g., a 6 was rotated to make a 9), the digit was counted as incorrect. According to measure developers (Shinn, 2019), AIMSweb™ does not provide specific information on the reliability and validity of the SSMM-Add/Sub measure. However, previous research has indicated there is validity in using CBM-M as a decision-making tool (Christ et al., 2008; Foegen et al., 2007; Shinn, 2004). In grades 3 through 5, alternative-form reliability estimates of CBM-M range from .72 to .93, with most estimates above .80. Internal consistency estimates of CBM-M are above .90 and validity estimates range from moderate (r = .35) to strong (r = .87; Foegen et al., 2007).

BEA Probes

Probes were created for progress monitoring across the eight sessions of the BEA, by dividing all addition and subtraction facts that include numerals 2-12 into 11 sets. Problems that included 0s and 1s were excluded so that the probes were appropriately challenging for students in both second and third grade (McCallum et al., 2006). Each probe was set up identically, with six problems per row and seven problems per column (Shinn et al., 1989). Problems were arranged in a stratified order, which ensured that they were repeated throughout the probe, but did not occur in the same order (Poncy et al. 2007; Poncy & Skinner, 2011). Probes were designed to be of equal difficulty, with one problem with 9 in each set, equal numbers of addition and subtraction problems in each probe, and equal numbers of problems across probes (Poncy et al., 2007). There were different problems on each probe, which ensured that there was no overlap of specific problems across the probes. The probes were then randomly assigned to conditions, so that overlap did not occur between the intervention conditions. Students were given 2-min to complete as many problems as possible and the probes were scored using the same procedures as SSMM-Add/Sub.

Consistent with recommendations from Daly et al. (1997), it was essential to ensure that the probes chosen for each student were equivalent in difficulty, so that differences observed across phases could be attributed to the intervention, not differences in probe difficulty. To determine probe equivalency, students completed all probes as part of the baseline phase in the BEA, and the difference from each participant's median score was determined for all probes. For each participant, the eight probes with the smallest difference from the median, and their corresponding intervention worksheet, were then randomly assigned to each intervention phase in the BEA (Szadokierski et al., 2017). The three probes with the greatest difference from the median score were not included. On average, each probe's difference from the median was: Probe 1 = -1.27, Probe 2 = -.9, Probe 3 = 1.18, Probe 4 = -3.77, Probe 5 = 1.09, Probe 6 = -1.93, Probe 7 = .07, Probe 8 = 2.23, Probe 9 = .55, Probe 10 = 3.18, and Probe 11 = 1.60. This indicates that, on average, Probes 8 and 10 were the easiest for participants while Probes 4 and 6 were the most difficult.

Social Validity Measures

As a measure of the social significance and importance of the intervention outcomes, social validity was measured using an adapted version of the Children's

Intervention Rating Profile (CIRP; Witt & Elliott, 1985) and an adapted version of the Intervention Rating Profile-15 (IRP-15; Witt & Martens, 1983). The CIRP and IRP-15 are the most commonly used measures employed in school psychology intervention research (Silva et al., 2020). The CIRP is a measure of the student's beliefs about an intervention that they have participated in. Two versions of the CIRP were created, one for each intervention. At the end of the intervention sessions, each CIRP was read to the student to determine their rating about each intervention that they participated in. The CIRP has an average coefficient alpha of .86 (Turco & Elliot, 1986). The IRP-15 is a measure of teachers' treatment acceptability. Two versions of the IRP-15 were created, one for each intervention. Teachers completed the IRP-15 after all student data was collected. The IRP-15's reported Cronbach's alpha is .98 (Witt & Martens, 1985).

Intervention

The two interventions were selected based on recommendations for acquisition and fluency interventions in computation (Codding et al., 2011; Barnett et al., 2004). The acquisition intervention was modeling with error correction and the fluency intervention was explicit timing with reward. Each intervention was designed to be completed in 10-min, so that time allocated to intervention was not a confounding factor in the study (Poncy et al., 2012). Consistent with recommendations from Daly et al. (1996), the same problems that were used for the progress monitoring probes were re-randomized and assigned to intervention worksheets. Having overlap between the intervention and assessment materials ensured that the progress monitoring materials were accurately capturing the effects of the treatment, enhancing the instructional validity of the progress monitoring probes (Daly et al., 1996; Daly et al., 1997).

Modeling with Error Correction

Modeling with error correction (M-EC) consisted of three steps. First, the examiner modeled problem solving on the instructional worksheet for 2-min, while the student followed along on their own worksheet (Everett et al., 2016; Mong & Mong, 2012; Mong et al., 2012). Next, the student practiced the worksheet for 2-min, while the examiner followed along and tracked student errors. Then, the interventionist provided error correction on any errors the student made, using an error correction procedure that followed a concrete-representational-abstract (CRA) sequence of instruction (Miller & Mercer, 1993). First, the interventionist modeled an incorrect item using a set of unifex cubes (e.g., “Six cubes, plus 2 cubes, equals 8 cubes”). Next, the interventionist modeled the problem using a number line from 0 to 25 (e.g., “We start on 6, and count up 2...7, 8. Six plus 2 equals 8”). Finally, the interventionist wrote the correct answer to the problem on the worksheet. The participant then modeled the problem using the unifex cubes and the number line and solved the problem correctly on their worksheet. Across all participants, 1.6 errors were corrected each intervention session (*SD* = 1.15). A BEA probe was then administered for 2-min for progress monitoring. Errors were not corrected during progress monitoring.

Explicit Timing with Reward

During explicit timing with reward (ET-R), the student practiced the intervention worksheet while being timed and was told to stop every 1-min. At the end of 1-min the student circled the last problem they completed (Rhymer et al., 2002). The examiner marked any errors and did not provide the correct answer. Students then took approximately 1-min to count the number of problems they completed correctly in 1-min, record this number on the side of their worksheet, and prepare for the next trial. They were then encouraged to beat their score on the next trial (Van Houten & Thompson, 1976; Rhymer et al., 1998). After completing these procedures 5 times, the interventionist asked the participant to select a reward from a prize box (e.g., gel pens, pencils, stickers, erasers; Duhon et al., 2004). The participant was told their goal (20% higher than their baseline rate; Soloman et al., 2018) and that they could earn the reward if they met or exceeded this goal. The participant then completed the progress monitoring probe for 2-min. The interventionist calculated the DC2M on the progress monitoring probe. If the participant met their goal, they were awarded the prize. If the student did not meet the goal, they were awarded a consolation prize of a sticker.

Study Procedures

The district had previously assigned each teacher in a grade to a "pod." All eligible participants were then assigned to one of five cohorts, based on the "pod" assignment of their teacher. For example, all study participants that were in Pods 1 and 2 were assigned to Cohort 1, Pods 3 and 4 were assigned to Cohort 2, etc. Cohort 1 had 17 participants, Cohort 2 had 20 participants, Cohort 3 and Cohort 4 each had 16 participants, and Cohort 5 had six participants. All baseline and instructional sessions were delivered individually by the first author, conducted in a grade-level workroom that was free of disruptions. Each student's intervention data collection occurred in seven sessions, implemented across two weeks. The data collection procedures were: (1) can't do/won't do assessment; (2) five or six BEA probes; (3) five or six BEA probes; (4) ET-R, M-EC; (5) ET-R, M-EC; (6) ET-R, M-EC; and (7) verification phase.

Can't Do/Won't Do Assessment

A can't do/won't do assessment (CDWD; VanDerHeyden & Witt, 2007) was conducted as the second gate of screening. The purpose of the CDWD was to verify the students' risk status and determine if students computation difficulties were due to a skill-only deficit (they cannot do it), a performance-only deficit (they will not do it), or a combined skill-performance deficit (they cannot and will not do it; VanDerHeyden & Witt, 2007). To conduct the CDWD, participants were administered a SSMM-Add/Sub, for 2-min. The participant's digits correct per 2-min (DC2M) was calculated, and the participant was shown a bag of rewards, which included pencils, gel pens, bookmarks, plastic rings, and erasers. The participants were told that if they could improve their DC2M by 20%, they could earn a prize (McKevett & Codding, 2019; Solomon et al., 2018). Prizes were provided if the student met or exceeded their goal. If students did not meet or exceed their goal, a sticker was provided as a consolation prize.

It was hypothesized that participants who could improve their score by 20%, but their resulting score was still below mastery (80 DC2M; Howell & Nolet, 1999), had a combined skill-performance deficit. It was hypothesized that participants who could not improve their score by 20% had a skill-only deficit (VanDerHeyden & Witt, 2007). It was hypothesized that participants who could meet or exceed the mastery criteria (80 DC2M) with the incentive had a performance-only deficit. Forty-four (59.5%) students were hypothesized to have a skill-only deficit, 30 (40.4%) students had a combined skill-performance deficit and no students had a performance-only deficit.

Baseline

To determine each participant's baseline performance, 11 probes were administered across two sessions, with five to six probes completed in each session

(Szadokierski et al., 2017). The median score from all 11 probes was used to determine each student's baseline rate and accuracy, which was consistent with recommendations for determining baseline academic performance, as the median is less sensitive to the impact of extreme scores (Shinn et al., 1989).

Brief Experimental Analysis Procedures

To determine which intervention would be most effective, a standard brief experimental analysis (BEA) was implemented with each participant. The standard BEA consisted of three, alternating phases of each intervention (Daly et al., 1997). Consistent with previous BEA research (Daly et al., 1999; McComas et al., 1996), the interventions were delivered from least to most intensive (ET-R, then M-EC).

An intervention was considered effective if it resulted in a score that was 20% or more than the participant's median baseline performance (McKevett & Codding, 2019; Soloman et al., 2018). Consistent with the methods of Szadokierski et al. (2017), if both interventions resulted in a score that was 20% or more than the student's median baseline performance, the best intervention was determined using the percentage of points exceeding the median (PEM; Ma, 2006). PEM requires the majority of data points (i.e., two of three data points) to exceed the median score of the other intervention. This ensured that the best intervention was more effective than baseline *and* the other intervention (Szadokierski et al., 2017). In the current study, both interventions resulted in a 20% increase in rate over baseline performance for 42 participants (56%); of these, ET-R was the most effective for 26 (62%) of students and M-EC was most effective for 16 (38%) of students. For 18 participants (24%), only ET-R resulted in a 20% increase in rate over baseline performance; and for 3 participants (5.3%), only M-EC resulted in a 20% increase in rate over baseline performance.

Inter-rater Reliability and Treatment Fidelity

Thirty percent of all measures were scored for inter-rater reliability by a doctoral-level school psychology graduate student, who had participated in graduate-level academic intervention and assessment courses. Training included a 30-min session on scoring for DC2M and accuracy. On the MCOMP, average inter-rater reliability was 99.6% (*range*, 95 to 100%). Average inter-rater reliability was 100% on the SSMM-Add/Sub and BEA probes.

Twenty percent of all intervention sessions implemented were randomly chosen to be recorded and evaluated to ensure treatment adherence. An independent observer, who was a doctoral-level special education student, analyzed treatment adherence by listening to the recorded intervention sessions implemented and completing an adherence checklist. Training was provided in a one-hour session, which included a description of the study, a description of each intervention, and how to score the treatment adherence protocol. Average treatment adherence was 96.8% (*range*, 95 to 100%) for M-EC sessions and 97% (*range*, 96 to 100%) for ET-R sessions.

Data Analysis

Analyses were conducted to classify students into intervention effectiveness groups and to conduct group comparisons, to answer each of the study's research questions.

Group Classification

Students were classified into one of three intervention effectiveness groups based on the results of the BEA, determined using PEM (Ma, 2006): (a) those for whom M-EC was most effective, (b) those for whom ET-R was most effective, and (c) those for whom neither intervention was effective.

Group Comparisons

The first research question was to determine if there are statistically significant differences in students' performance on the measures included in the gated screening framework (STAR Math, MCOMP, SSMM-Add/Sub), between those for whom M-EC was most effective and those for whom ET-R was most effective. Because the second grade and third grade version of the MCOMP has different problems, with different total points possible, all MCOMP total scores were converted to z-scores, to allow for analyses to be conducted across both grades (Field, 2005). A multi-variate analysis of variance (MANOVA) was conducted to evaluate the mean differences in STAR Math, MCOMP, SSMM-Add/Sub rate and SSMM-Add/Sub accuracy. Planned comparisons were then conducted to evaluate the two pairwise comparisons of means within the MANOVA. Because the two groups (M-EC and ET-R) had different participant sizes, Hedge's *g* was used to compute effect sizes between the mean performance of the M-EC group and the mean performance of the ET-R group (Hedges, 1982). Students for whom neither intervention was effective were not included in this analysis (Szadokierski et al., 2017).

The second research question was to determine which variables in the gated screening framework (STAR Math, MCOMP, results of the CDWD assessment, SSMM-Add/Sub rate, and SSMM-Add/Sub accuracy) best predicted which intervention was the most effective in the BEA. Logistical regression models were used, rather than linear models, because the resulting effectiveness data was dichotomous (M-EC or ET-R was the best intervention). Students for whom neither intervention was effective were not included in this analysis (Szadokierski et al., 2017).

The final research question was an exploratory analysis, to determine the cut scores, or range of cut scores, on the measures that best predict which intervention will be most effective. Receiver operating characteristic (ROC) curves were used to evaluate the prediction quality of the single-variable models. ROC curves provide information about the sensitivity (rate of true positives) and specificity (rate of true negative) at different points in the model (Fawcett, 2006). Additionally, the area under each curve (AUC) was also examined as an indicator of overall accuracy of the model. The further away from .5 (indicates a 50/50 chance) and the closer it is to one, the better the measure differentiates between students for whom an intervention is effective from those for whom it is not (Metz, 1978).

Results

Box plots were examined to determine if there were any outliers in the distribution of the data and the distribution of the data was evaluated by examining the z-scores of skewness and kurtosis (Field, 2005). The z-scores for skewness were: STAR Math = -2.26, MCOMP = 2.67, SSMM-rate = 0.56, and SSMM-accuracy = 8.01. The z-scores for kurtosis were: STAR Math = 1.40, MCOMP = 1.24, SSMM-rate = 1.43, and SSMM-accuracy = 17.51. STAR Math, MCOMP and SSMM-rate were acceptably distributed (Field, 2005). The distribution of SSMM-accuracy was significantly left skewed, so this variable was removed from all planned statistical analyses. This decision was also supported by previous empirical evidence which suggests that in math accuracy is not as important of an indicator as rate (Burns et al., 2006),

Descriptive Analyses

STAR Math scores ranged from 211 to 632, with a mean of 479.85 (*SD* = 80.65). The mean MCOMP score was at the 26th percentile for second grade (*M* = 24.37, *SD* = 5.85, *range* = 7 to 31) and at the 26th percentile for third grade (*M* = 30.34, *SD* = 8.34, *range* = 6 to 40). The mean SSMM-rate was within the frustrational range (*M* = 25.08 DC2M, *SD* = 8.15 DC2M, *range* = 2 to 50 DC2M) and the mean SSMM-accuracy was within the instructional range (*M* = 93.48% *SD* = 8.09, *range* = 47.2% to 100%; Howell & Nolet, 1999). STAR Math was moderately correlated with MCOMP (*r* = .58, *p* = .00) and SSMM-rate (*r* = .53, *p* = .00). MCOMP was also moderately correlated with SSMM-rate (*r* = .55, *p* = .00).

M-EC was most effective for 19 students (25.3%) and ET-R was most effective for 44 students (58.7%). See Appendix A for an example of a participant for whom all of the effective intervention's data points fell above the median of the less effective intervention and Appendix B for an example of a participant for whom 66% of the effective intervention's data points fell above the other intervention's median. Neither intervention was differentially effective for 11 students (14.7%), which is shown in Appendix C. Among the students for whom M-EC was most effective, all three data points exceeded the median of ET-R for 8 students (42.1%), and two of the three data points exceeded the median of ET-R for 11 students (57.9%). Among the students for whom ET-R was more effective, all three data points exceeded the median of M-EC for 27 students (61.4%), and two of the three data points exceeded the median of M-EC for

17 students (38.6%). The verification phase was consistent with the best intervention from the BEA for 82.5% (*n* =52) of the participants in which a differentially effective intervention could be identified.

Chi-squared analyses were conducted to determine if there was a significant association between grade, gender, special education status, or results of the CDWD assessment across intervention groups. The chi-squared analyses found no significant association between grade and intervention group, $\chi^2(df = 2) = 0.57$, $p = .75$, gender and intervention group, $\chi^2(df = 2) = 0.81$, $p = .67$, special education status and intervention group, $\chi^2(df = 2) = 0.57$, $p = .75$, or results of the CDWD assessment and the intervention group, $\chi^2(df = 2) = 0.80$, $p = .67$. These results indicated that there was an equal distribution across the intervention groups of grade, gender, special education status, and results of the CDWD assessment.

Significant Differences in Preintervention Performance

A MANOVA was conducted to determine whether there were differences between intervention groups for STAR Math, MCOMP, or SSMM-rate. Because multiple comparisons were conducted, a Bonferonni corrected alpha level of .013 was used (Field, 2005). A significant effect was found between intervention groups for SSMM-rate, $F(1, 61) = 8.2$, $p = .006$, but not for STAR Math, $F(1, 61) = 6.4$, $p = .014$, or MCOMP, $F(1, 61) = 1.9$, $p = 0.169$. A planned pairwise comparison between the mean SSMM-rate for the two intervention groups indicated a significant and large effect, $t(61) = 3.03$, $p = .004$, $g = 0.84$.

Predicting the Best Intervention

Next, single variable logistic regression models were used to determine how well STAR Math, MCOMP, results of the CDWD assessment, and SSMM-rate predicted which intervention group would be most effective. Table 1 shows the results of the single variable logistic regression models. Results indicate that STAR Math and SSMM-rate were significant predictors of intervention effectiveness group. In logistic regression, the pseudo R^2 values cannot be interpreted in the same way as in linear regression, but they do provide a general comparative indicator of model fit (Field, 2005). Overall, the pseudo R^2 values and chi-square values indicate that STAR Math and SSMM-rate improve the prediction of whether M-EC or ET-R would be most successful, over a constant-only, null model. Table 1 also provides the percentage of cases that were correctly predicted by each single variable model. STAR Math (75.4%) and SSMM-rate (71.4%) both predicted the greatest number of cases.

Exploratory Cut Score Analysis

ROC analyses were conducted using the variables that were significantly different across groups, which was only SSMM-rate. The ROC analysis resulted in an AUC of .72 for the model that used SSMM-rate as the predictor. This is above the level that would be expected by chance (.50) and within the acceptable range (.70 to .80; Mandrekar, 2010). Thus, the results of the ROC analysis indicate that SSMM-rate was able to differentiate between students who responded to each intervention, from those who did not. Using the ROC curves, the cut point for each model that optimizes sensitivity (the rate of true positives) and specificity (the rate of true negatives) was determined. The optimal cut score for M-EC was a rate of 23.5 DC2M; students who scored below 23.5 were more likely to respond best to M-EC. The optimal cut score for ET-R was 27.5 DC2M;

students who scored above 27.5 DC2M were more likely to respond best to ET-R. The two cut scores can be averaged to provide a precise cut score (Szadokierski et al., 2017), with a resulting score of 25.5 DC2M. This score was rounded up, as it is impossible to earn a half-point for a correct digit, resulting in a cut score of 26 DC2M. Thirty-nine participants (52.7%) scored below this cut point, while 35 participants (47.3%) scored above this cut point. The sensitivity and specificity of the SSMM-rate metric in identifying students who would respond best to M-EC was 68.2% and 42.1%, respectively. The sensitivity and specificity of the pretest-rate metric in identifying students who would respond best to ET-R was 73.7% and 45.5%, respectively. The National Center on Intensive Intervention's (NCII's) tools chart rates a screening tool as having convincing evidence when it has a sensitivity level rate of 70% or higher and a specificity rate of at least 80%. Based on this criteria, the results suggest that the SSMM-rate metric may be an acceptable measure in identifying true positives but may be an unacceptable metric in identifying true negatives.

Social Validity Scales

Appendices D and E show the results of the adapted CIRP and IRP-15, respectively. Results of the adapted CIRP indicate that overall, students found M-EC and ET-R to be equally acceptable. Average acceptability was 33.9 (*SD* = 9.4, *range* = 20 to 42) for M-EC and 32.6 (*SD* = 9.8, *range* = 24 to 42) for ET-R. Results of the IRP-15 indicated that overall, teachers found both interventions to be acceptable; however, teachers found M-EC (*M* = 72.6, *SD* = 6.4, *range* = 59 to 85) to be more acceptable for use in improving students' computation skills than ET-R (*M* = 60.8, *SD* = 9.8, *range* = 46 to 72).

Discussion

The purpose of the current study was to evaluate a gated screening framework in math that could be used for instructional planning for students who are identified as at-risk in whole number proficiency. Unlike in reading, where a single, 1-min measure of oral reading fluency can be used for both universal screening and intervention planning (Szadokierski et al., 2017), math is a multifaceted subject which requires a different approach. The current recommendations for screening in math do not lend themselves to instructional planning, and recommendations for instructional planning (i.e., BEAs and diagnostic assessment) are time intensive procedures for educators to complete. Therefore, there is a need to investigate ways to go from screening to effective intervention that will rely on fewer resources and less time.

The current study evaluated a gated screening framework that included STAR Math, MCOMP, and a CDWD assessment using SSMM-Add/Sub. A standard BEA was used to evaluate which of two interventions, M-EC or ET-R, was most effective for each student. Group analyses were planned to include both SSMM-rate and SSMM-accuracy; however, the distribution of the SSMM-accuracy data had little variability and was significantly left skewed. Therefore, SSMM-accuracy was removed from all planned analyses. Statistically significant differences between intervention groups were yielded for SSMM-rate but not on the broader universal screening measures of STAR Math or MCOMP. Single variable logistic regression models indicated that Star Math and SSMM-rate were able to predict which intervention was most effective, while MCOMP and the results of the CDWD assessment were not able to predict which intervention was most effective. A ROC analysis was used to determine the optimal cut score for SSMM-rate, which was 26 DC2M (13 DCPM). Overall, these results support other research (VanNorman et al., 2018, Vaughn & Fletcher, 2012), which suggests that given broader universal screening measures in math may not provide sufficient information to guide instructional planning, it may be necessary to include a subskill mastery measure within a gated screening framework in math. These measures, such as a measure of whole number proficiency skills, can provide more specific information on the students' current skill level, to be able to accurately match students to interventions.

A Gated Screening Framework for Math

Consensus among experts appears to be converging around the idea that gated screening approaches might be the most appropriate way to minimize false-negative and false-positive decisions (VanDerHeyden, 2013; VanDerHeyden et al., 2017; Compton et al., 2010; Compton et al., 2012; Nelson et al., 2016). A gated screening approach may be particularly useful in math given the multi-faceted nature of the subject and the hierarchical nature of skills (Compton et al., 2010; Nelson et al., 2016). A gated screening framework can also provide information to identify students who are at-risk and determine appropriate instructional procedures (Van Norman et al., 2018). There are several options for gated screening that have been tested in the literature; the current study converges with recommendations that suggest that the first step may be to simply identify students who are not proficient (Compton et al., 2010; Compton et al., 2012, Fuchs et al., 2012; Nelson et al., 2016; VanDerHeyden et al., 2017). Then, when additional information is gathered in subsequent gates, it may be useful to include measures of subskill mastery in order to promote better instructional match to students' stage of skill development, particularly when whole number computation is the instructional target (Hosp & Ardoin, 2008; Shapiro, 2011; Van Norman et al., 2018; & Vaughn & Fletcher, 2012).

The current study utilized MCOMP in the first gate, because this measure has been demonstrated to be a strong predictor of students' math achievement levels (VanDerHeyden et al., 2017). The STAR Math, a computer adaptive test with adequate sensitivity and sensitivity for identifying risk status was also used to verify students' risk status. This study expands the literature on gated screening by examining if various screening tools could predict the most effective of two skill-based interventions (M-EC or ET-R) when effectiveness was determined using a standard BEA. Neither STAR Math nor the MCOMP reliably differentiated between students who would benefit from an acquisition or a fluency intervention. STAR Math is a broad computer adaptive test designed to estimate skills 32 domains, including numeration concepts, computation, word problems, estimation, data analysis and statistics, geometry, measurement, and algebra, and has been shown to predict student growth well in third through fifth grades (Renaissance Learning, 2016; Shapiro et al., 2015). STAR Math has also been shown to be accurate in predicting student performance on end-of-year state standards (Renaissance Learning, 2016). While STAR Math is a strong predictor of students' risk status and end-of-year standards, it is possible that STAR Math was too broad of a measure to accurately differentiate between students whose whole number proficiency skills were within the acquisition level or within the fluency level. With respect to MCOMP, although MCOMP is a more proximal measure, given that all the items focus on arithmetic, the measure still represents a wide sampling of arithmetic skills from the curriculum for each grade level and was not a direct measure of the intervention outcomes (Foegen et al., 2007). Hintze et al. (2002) found that measures that represent complex curriculum sampling, such as MCOMP, require that students take multiple forms to reliably gauge math performance levels. Therefore, it may have been necessary to gather information on multiple forms of MCOMP to reliably estimate each student's performance level.

The CDWD assessment was included as part of the gated screening process in order to ensure that all students who were identified as at-risk for mathematics failure would need a skill-based intervention. The results of the CDWD assessment were not able to predict which skill-based intervention would be most effective for each student. Consistent with the study by Ardoin and colleagues (2005), the CDWD assessment was able to verify that all students who were found to be eligible were an appropriate match for a skill-based intervention, as none of the students were identified as having a performance-only deficit. Forty-four (59.5%) students had a skill-only deficit, and 30 (40.4%) students had a combined skill-performance deficit.

Cut-scores for Subskill Mastery Measures

The current study found that students who score below 26 DC2M (13 DCPM) benefited most from the acquisition intervention whereas those students who scored above the criterion benefited most from the fluency intervention. This finding converges with the only other empirical study to determine the curriculum-based assessment criteria that aligns with the instructional hierarchy (Burns et al., 2006), which demonstrated that that for second and third grade, students who score below 14 digits correct per minute (DCPM) are within an acquisition stage of learning. Prior recommendations by Deno and Mirkin (1977) and Howell and Nolet (1999), suggested cutoff scores of 10 DCPM and 20

DCPM, respectively. However, the current findings are compelling because the cut scores were empirically derived, are similar to those derived by Burns et al. (2006). This finding also supports the call from Jenkins et al. (2007) for more studies on screening measures that cross-validate cut points with new samples of students, further supporting the cut points that have been empirically derived. Burns et al. (2006) was conducted in a suburban setting with a larger sample (434 students), 74% of which identified as Caucasian (not Hispanic), 17% as Hispanic or Latino, 6% as African American, 3% as Asian American, and 1% as Native American. Six percent of participants in Burns et al. (2006) were eligible for special education. The current student provides additional generalizability to the findings by occurring in a rural setting, including Native American students (39% of the study sample) and special education students (16.9% of the study sample), and collecting data on the SES status of participating students (70.1% of participants were eligible for free or reduced price lunch).

Findings from the ROC analysis were that the SSMM-rate metric may be an acceptable measure for identifying true positives but may not an acceptable metric for identifying true negatives. Sensitivity for identifying students who would respond best to M-EC was 68.2% and for identifying students who would respond best to ET-R was 73.7%. Specificity for identifying students who would respond best to M-EC was 42.1% and for identifying students who would respond best to ET-R was 45.5%. Using criteria from NCII's tools chart, which indicates that a tool is acceptable when it has a sensitivity level rate of at least 70% and a specificity rate of at least 80%, the SSMM-rate metric meets criteria for identifying true positives but is not for identifying true negatives. According to the NCII's tools chart, there are few mathematics screening measures that meet this criteria, with just three out of 15 elementary mathematics screening measures rated as meeting the full criterion for the winter screening period. In the current study, high false negative rates meant that it was possible that the incorrect intervention was identified as being most effective. If this intervention was delivered over time, it is possible that time and resources would be wasted on a mis-matched intervention (Fuchs et al., 2007). However, through progress monitoring and the application of data-based decision making, educators can make timely changes to intervention that could mitigate the potential for a false negative (Fuchs et al., 2007).

Acceptability of Brief Interventions

Within this study a brief experimental analysis was used to determine which intervention would be most effective for each student. This finding then served as the basis from which to determine which screening tool or tools would accurately predict instructional intervention match. The interventions that we created to generate skill accuracy or skill fluency options were designed so as to match theoretical recommendations from the instructional hierarchy (Haring & Eaton, 1978) and recommendations from previous empirical studies that examined the effects of acquisition and fluency interventions in whole number computation (Codding et al., 2007; Burns et al., 2010). As such we were interested in obtaining descriptive information on social validity of these options. Social validity is important to consider because if teachers or students do not find a chosen intervention to be acceptable, it is unlikely that they will engage in the intervention in the future (Kratochwill et al., 2018). Results of the current study indicated that students found both interventions to be equally acceptable. However, teachers found M-EC to be more acceptable than ET-R, indicating that M-EC would be more likely to be used in their classroom. This may be due to the fact that many teachers believe that teaching conceptual understanding should take precedence over building procedural knowledge and that timed tasks can increase math anxiety (VanDerHeyden & Codding, 2020). There is also a well-known gap in fluency building in schools, where math instruction and associated curricular materials often do not include adequate opportunities to respond, through repeated practice (NCTM, 2014; NMAP, 2008; Kang, 2016). However, using repeated, timed practice opportunities has been demonstrated to be a high-quality fluency building instructional procedure, particularly when it incorporates performance feedback and contingent reinforcement, and using timed practice or timed assessments is not associated with increased anxiety (Codding et al., 2007; 2011; VanDerHeyden & Codding, 2020).

Implications for Research and Practice

The results of this study provide additional information on the use of a gated screening framework for instructional planning in whole number proficiency, and also provide direction for future research and practical application. First, there is a need for additional research on gated screening in mathematics using a variety of screening measures that are commonly implemented in schools. The current methods could be replicated with other commonly available measures, such as those from FastBridge or iReady. Second, there is a need for additional research on the classification accuracy of mathematics screening measures, both for identifying student as at-risk and for instructional planning. Without established criterion for classification accuracy, schools may be unclear about which measures will best meet their needs. In practice, this study further supports the use of a gated screening framework in math.

Educators should follow the recommendations from VanDerHeyden et al. (2017) and Van Norman et al. (2018), to use the previous year's test scores when available or a strong initial screener to identify students as at risk. Then, educators could use CDWD assessment to determine if a student's difficulties are due to a performance-only deficit, and if the student would be an appropriate candidate for a skill-based intervention (Ardoin et al., 2005). Finally, a subskill mastery measure may be useful to more accurately identify the student's level of learning, applying the instructional hierarchy. In the area of whole number proficiency, educators can use the criterion of 13-14 digits correct as a cut point to differentiate between students who would most likely benefit from an acquisition-based or fluency-based intervention.

Limitations and Future Directions

The findings of this paper should be considered within the context of a few limitations. A limitation of the current study was the sample size. The a priori power analysis indicated that 65 to 87 participants were required to detect a moderate to large effect. The current study was conducted with 75 students, meaning that there was sufficient power to detect a large effect. However, for the data analyses, the students for whom neither intervention was differentially effective were excluded, which lowered the number of students included in these analyses to 63. Consequently, it is possible that the statistical analyses were underpowered to detect a large effect across all three variables (MCOMP, STAR Math and SSMM-rate). This could have impacted the finding that STAR Math was unable to differentiate between the two intervention groups, as the significance level of Star Math was .014 in this analysis, whereas the Bonferroni corrected alpha was .013. That said, the sample size was greater than that in similar prior research conducted by Szadokierski et al. (2017), which recruited 43 participants, and is the only study that has been conducted with a similar design. Additionally, although the relatively diverse sample is a strength of this study in comparison to other similar studies, a limitation of the current study sample was that the primary language of all students in the sample was English. Future replications of this study with diverse samples that also include students whose primary language is not English are needed.

Conclusion

This study demonstrated that a gated screening framework can be used to match students who are identified as at risk for mathematics failure to an effective whole number computation intervention. Among a sample of students who were identified as at risk, those who completed less than 26 DC2M on the SSMM-Add/Sub benefitted most from an acquisition intervention, while those who completed more than 26 DC2M benefitted most from a fluency intervention. These guidelines can be used to guide educators when using the problem solving model address students' intervention in an accurate and efficient manner needs once screening as identified those students as needing additional supports.

Table 1

Results of the Binary Logistic Regression

	β	SE	Pseudo R^2	χ^2	Percent Correct
Model 1					
STAR Math	-.01*	.00	.13	6.04*	75.4%
Model 2					
MCOMP	-.38	0.28	.04	1.86	69.8%
Model 3					
CDWD	-.454	.56	.02	0.67	69.8%
Model 4					
SSMM-Rate	-0.11*	0.04	.19	8.89*	71.4%

Note. * $p < .05$. CDWD = Can't Do/Won't Do Assessment; MCOMP = AIMSweb™ Math Computation.

Chapter 4

The results of these two studies provide evidence for two methods of instructional planning, BEAs and a gated screening framework, for students who are identified as needing supplemental intervention in whole number proficiency. The first study was a systematic literature review of 16 studies, with 67 participants, that used a BEA to determine the most effective mathematics intervention for students. This study provided a better understanding of how researchers and practitioners can use a BEA to determine an effective mathematics intervention; however, highlighted many technical limitations of using a BEA to identify a mathematics intervention, and several directions for future research.

The second study evaluated a gated screening framework in math that could be used for instructional planning for students who were identified as at-risk in whole number proficiency. The gated screening framework included STAR Math, MCOMP, and a CDWD assessment using SSMM-Add/Sub. A standard BEA was then used to evaluate which of two interventions, modeling with error correction or explicit timing with reward, was most effective for each student. Statistically significant differences were found between intervention groups for SSMM-rate but not on the broader universal screening measures of STAR Math or MCOMP. Star Math and SSMM-rate were able to predict which intervention was most effective, while MCOMP and the results of the CDWD assessment were not able to predict which intervention was most effective. A ROC analysis was used to determine the optimal cut score for SSMM-rate, which was 26 DC2M (13 DCPM). This meant that students who scored below 13 DCPM were likely to respond best to M-EC, while those who scored above 13 DCPM were likely to respond best to ET-R. Overall, these results support other research (VanNorman et al., 2018, Vaughn & Fletcher, 2012), which suggests that it may be necessary to include a subskill mastery measure within a gated screening framework in math, to be able to accurately match students to interventions. Additionally, when considered together, both studies demonstrate that BEAs may be used to identify an effective intervention for whole number proficiency, when considerations are made for appropriate measurement of the outcome variable; however, a gated screening framework may be a more time efficient procedure for educators to implement to accurately match a student to an effective intervention for whole number proficiency.

Study 1: Systematic Review of Mathematics BEA Studies

In Study 1, a systematic review was used to locate all empirical studies that have used a BEA to determine the most effective math intervention for students. Several limitations of the math BEA literature were noted in this study, which are important for both practitioners and future researchers to consider. The first limitation was that the ethnicity of the study participants was limited. Study participants were primarily white or black English-speaking students, with just five Hispanic students included. Additionally, only four students were identified as eligible for special education, under the category of learning disability. This limits the extent to which the results of these studies can be applied to students from a more diverse population, has important implications for practitioners who may attempt to use a BEA within their school psychology practice, and requires that researchers be more intentional about including a diverse set of students in their research.

The second limitation was that all BEA studies targeted whole number computation, with digits correct per minute serving as the primary dependent variable. To measure intervention outcomes, researchers developed their own probes. Four studies (Codding et al., 2019; Hofstader-Duke & Daly, 2014; Kleinert, 2017; Silva, 2017) developed probe sets and assigned problems to treatment conditions. The remaining studies did not assign problems to treatment conditions, and instead developed progress monitoring probes using a random generator or the Math Worksheet Generator from interventioncentral.org or mathaids.com. For these studies, there was a lack of overlap between assessment and intervention materials, which makes it difficult to determine if the probes were capturing the effects of either the treatment or repeated exposure to the same problems across probes and conditions (Daly et al., 1996; Daly et al., 1997). Additionally, Strait et al. (2015) demonstrated that randomized probes from online worksheet generators do not meet the reliability thresholds that are required for progress monitoring (0.80), which increases the measurement error in student outcomes. These concerns with measurement are important for researchers to consider and remediate in future studies. It is important for researchers to remember that when conducting BEAs in math, the dependent measures should: 1) match the intervention content; 2) be equally difficult, so that difficulty across probes does not explain differences in student performance; 3) problems should be assigned to conditions, to reduce the potential for overlap in learning across conditions; and 4) probes should be designed in ways that are consistent with research-based recommendations for progress monitoring research, such as being stratified (rather than randomized), to ensure that problems do not occur in the same order throughout the probe (Poncy et al., 2007; Poncy & Skinner, 2011).

Finally, an assessment of study quality indicated that the majority of previous BEA studies in math did not consider social validity. Social validity is an important component of assessment and intervention research, because if the social validity of either the assessment practice or an intervention is low, neither are likely to be implemented in the future (Kratochwill et al., 2018). Future researchers may consider including measures such as the Children's Intervention Rating Profile (CIRP; Witt & Elliot, 1985) and the Intervention Rating Profile-15 (IRP-15), which are the most common measures in school psychology intervention research, to evaluate the social validity of the interventions included in the BEA (Silva et al., 2020).

Study 2: Gated Screening Framework in Math

The second study evaluated a gated screening framework for instructional planning in math, using STAR Math, MCOMP, and a CDWD assessment using SSMM-Add/Sub. A MANOVA was conducted to determine if there were significant differences across the intervention groups on the measures included in the gated screening framework. A significant and large effect was found between intervention groups for SSMM-rate, but not for STAR Math or MCOMP. Single variable logistical regression models were also examined to determine how well STAR Math, MCOMP, results of the CDWD assessment and SSMM-rate predicted which intervention group would be most effective. Results indicated that STAR Math and SSMM-rate were significant predictors of intervention effectiveness group. A cut score analysis was conducted to explore which cut score for SSMM-rate could be used to accurately assign students to an intervention. The cut score analysis indicated that those who earned below 26 DC2M (13 DCPM) responded best to M-EC and those who earned above 26 DC2M responded best to ET-R.

Determining the Best Intervention

To determine which intervention, M-EC or ET-R, was most effective for each student, a standard BEA was used. This study attempted to remediate many of the technical limitations that were identified in Study 1. First, this study attempted to remediate the issue of inadequate dependent measures by creating stratified probes and assigning problems to probes, probes to intervention conditions and the same problems to intervention worksheets. This ensured that we were assessing the effects of the intervention, rather than repeated exposure to problems, or carryover in skill development across intervention conditions. Next, we included measures of social validity to determine if students and/or teachers found M-EC or ET-R to be acceptable. Results indicated that students found each intervention to be equally acceptable, while teachers found M-EC to be more acceptable than ET-R. This may be due to the fact that many teachers believe that conceptual understanding should take precedence over procedural knowledge and that timed tasks can cause mathematics anxiety (VanDerHeyden & Codding, 2020). However, research does not support these myths, and rather conceptual and procedural learning should occur at the same time and using repeated and timed practice opportunities has been found to be a high-quality and evidence-based instructional strategy to build proficiency (VanDerHeyden & Codding, 2020). Finally, this study attempted to increase the diversity of study participants in the math BEA research, by being conducted in a rural setting, including Native American students (39% of the study sample), special education students (16.9% of the study sample) and including information on the SES status of participants (70.1% of participants were eligible for free or reduced priced lunch). However, a limitation was that this study did not include students whose primary language was different from English.

Overall, these two studies support the use of BEAs and a gated screening framework for instructional planning in whole number computation. To conduct a BEA in math, educators may consider using an abridged-BEA design for efficiency, or a standard-BEA design to increase their confidence in the outcomes of the BEA. However, BEAs have only been used for whole number computation, and with students who are white, black, and Native American, and whose primary language is English. Additionally, when conducting universal screening for math, educators should use a gated screening framework to not only identify students who are at risk, but to also match students who need supplemental intervention in whole number proficiency to an intervention that is going to meet their needs. A gated screening framework should include a strong universal screener in grades K-3, year-end test scores in grades 4-5 and high local cut scores to reduce the rate of false negatives (Van Norman et al., 2018). Educators can also consider using a CDWD assessment to verify that students are appropriate for a skill-based intervention (Ardoin et al., 2005). Finally, educators should include a SSMM of the intervention skill to adequately determine if a student would be most likely to benefit from an acquisition-based or fluency-based intervention.

Future Research

Based on these two studies, several questions emerge for future research. First, more research is needed on how to use both of these formats for different concepts in mathematics. While it is essential for elementary students to master whole number proficiency, math is multifaceted and it would be important for researchers to explore these methods for matching students to interventions in other mathematical concepts, such as early numeracy skills and rational number knowledge. There is also a need to evaluate the gated screening framework that was evaluated in Study 2 with more students. Unfortunately, it is possible that this study was underpowered, which may have affected the results for STAR Math. This is important as schools attempt to implement MTSS and RtI in mathematics and look for ways to accurately match students to interventions in the most time and resource efficient manner. If schools could detect difference with a single measure, it would mean that they would not need to conduct additional gates in their screening model. However, at this time, this study supports the use of a gated screening framework for instructional planning. Finally, all studies located in Study 1, and Study 2, were conducted by researchers. Future researchers should explore if outcomes are the same when implemented by natural implementers, such as practicing school psychologists, interventionists, or special educators. Doing so will further increase the social validity of both methods for instructional planning.

Conclusion

This dissertation aimed to increase both practitioners' and researchers' understanding of how to accurately and efficiently match students to a whole number proficiency intervention that will meet their needs. This is essential as more schools attempt to implement MTSS and RtI in math, as schools cannot afford to waste time and resources by implementing interventions that are unlikely to work for students. Findings from these studies indicate that more research is needed to evaluate the technical properties of using BEAs to identify an effective intervention. Additionally, a gated screening framework can be used to identify students who are at risk and match them to an effective intervention for whole number proficiency. Students who score below 26 DC2M on a SSMM are likely to benefit the most from an acquisition-based intervention, while those who can complete more than 26 DC2M are likely to benefit the most from a fluency-based intervention.

References

References marked with an asterisk indicate studies included in the synthesis.

AIMSweb Technical Manual (2012). Retrieved from: https://aimsweb.pearson.com/.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C. & Wagner, R. K. (2014). To wait in Tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children, 81*(1), 11-27.

Al Otaiba, S. & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23*(5), 300-316.Andersen, M. A., Daly, E. J., & Young, N. D. (2013). Examination of one-trial brief experimental analysis to identify reading fluency interventions. *Psychology in the Schools, 50*(4), 403-414.

Ardoin, S. P. & Daly, E. J. (2007). Introduction to the special series: Close encounters of the instructional kind – how the instructional hierarchy is shaping instructional research 30 years later. *Journal of Behavioral Education, 16,* 1-6.

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23*(4), 362-380.

*Atbasi, S. & Sanir, H. (2018). Comparison of performance-based techniques to increase addition fluency: A brief experimental analysis study. *Education and Science, 43*(195), 241-252.

Backman, C. L., Harris, S. R., Chisholm, J. M., Monette, A. D. (1997). Single-subject research in rehabilitation: A review of studies using AB, withdrawal, multiple baseline, and alternating treatments design. *Archives of Physical Medicine and Rehabilitation, 78*(1), 1145-1153.

Barlow, D. H. & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*(2), 199-210.

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, F. E. (2004). Response to intervention: Empirically-based special service decisions from single-case designs of increasing and decreasing intensity. *The Journal of Special Education, 38*, 66-79.

Bond, L. A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation, 5*(2), 1-3.

Burns, M. K., VanDerHeyden, A. M., Zaslofsky, A. F. (2014). Best practices in delivering intensive academic interventions with a skill-by-treatment interaction. *Best Practices in School Psychology VI.* Bethesda MD: National Association of School Psychologists.

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review, 39*(1), 69-83.

Burns, M. K. & Coolong-Chaffin, M. (2006). Response to intervention: The role of and effect on school psychology. *School Psychology Forum: Research in Practice, 1*(1), 3-15.

Burns, M. K., Ganuza, Z. M. & London, R. M. (2009). Brief experimental analysis of written letter formation: Single-case demonstration. *Journal of Behavioral Education, 18*, 20-34.

Burns, M. K., Jimerson, S. R., VanDerHeyden, A. M., & Deno, S. L. (2015). Toward a unified response-to-intervention model: Multi-tiered systems of support. In *Handbook of Response to Intervention*, Springer Science+Business Media: New York.

Burns, M. K., Taylor, C. N., Warmbold-Brann, K. L., Preast, J. L., Hosp, J. L., & Ford, J. W. (2017). Empirical synthesis of the effect of standard error of measurement on decisions made within brief experimental analyses of reading fluency. *Psychology in the Schools, 54*(6), 640-654.

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. L. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review*, *35*(3), 401-418.

Burns, M. K. & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review, 37*(1), 126-136.

Carnine, D., Silbert, J., & Kameenui, E. J. (1990). *Direct instruction reading* (2nd Ed.) Merrill: Columbus, OH.

*Carson, P. M. & Eckert, T. L. (2003). An experimental analysis of mathematics instructional components: Examining the effects of student-selected versus empirically-selected interventions. *Journal of Behavioral Education, 12*(1), 35-54.

Christ, T. J. & Aranas, Y. A. (2008). Best practices in problem analysis. *Best Practices in School Psychology.* Bethesda, MD: National Association of School Psychologists.

Christ, T. J., Burns, M. K. & Ysseldyke, J. E. (2005). Conceptual confusion within response-to-intervention vernacular: Clarifying meaningful differences. *NASP Communique, 34*(3), 1-8.

Christ, T. J., Scullin, S., Tolbize, A., & Jiban, C. L. (2008). Implication of recent research: Curriculum-based measurement of math computation. *Assessment for Effective Intervention, 33*(4), 198-205.

Christ, T. J. & Vining, O. (2006). Curriculum-based measurement procedures to develop multiple-skill mathematics computation probes: Evaluation of random and stratified stimulus-set arrangements. *School Psychology Review, 35*(3), 387-400

*Clark, C. R. (2013). *Using brief experimental analyses to identify effective math interventions for early elementary students.* Doctoral dissertation. Retrieved from PsychInfo.

*Codding, R. S., Baglici, S., Gottesman, D., Johnson, M., Kert, A. S., & Lebouf, P. (2009). Selecting intervention strategies: Using brief experimental analysis for mathematics problems. *Journal of Applied School Psychology, 25*(2), 146-168.

Codding, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice*, *26*(1), 36-47.

Codding, R. S., Shiyko, M., Russo, M., Birch, S., Fanning, E., & Jaspen, D. (2007). Comparing mathematics interventions: Does initial level of fluency predict intervention effectiveness? *Journal of School Psychology*, *45*(6), 603-617.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquiero, L. A., […] & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, 102(2), 327-340.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*, 394-409.

Compton, D. L., Gilbert, J. R., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho. E., […] Bouton, B. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities, 45*(3), 204-216.

Connor, C. M., Spender, M., Day, S. L., Giuliani, S., Ingebrand, S. W., McLean, L., & Morrison, F. J. (2014). Capturing the complexity: Content, type and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes. *Journal of Educational Psychology*, 106(3), 762-778.

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology,* 66, 4-10.

Daly, E. J., Bonfiglio, C. M., Mattson, T., Persampieri, M., & Foreman-Yates, K. (2006). Refining the experimental analysis of academic skills deficits: Part II. Use of brief experimental analysis to evaluate reading fluency treatments. *Journal of Applied Behavior Analysis, 39*, 323-331.

Daly, E. J., Martens, B. K., Hamler, K. R., Dool, E. J. & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32*(1), 83-94.

Daly, E. J., Martens, B. K., Kilmer, A., & Massie, D. R. (1996). The effects of instructional match and content overlap on generalized reading performance. *Journal of Applied Behavior Analysis, 29*(4), 507-518.

Daly, E. J., Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*(4), 554-574.

Dennis, M. S. (2015). Effects of tier 2 and tier 3 mathematics interventions for second graders with mathematics difficulties. *Learning Disabilities Research & Practice,* 30(1), 29-42.

Deno, S. L. (2015). Problem-solving assessment. In R. Brown-Chidsey & K. J. Andren (Eds)., *Assessment for Intervention: A Problem-Solving Approach*, (pp. 10-39). New York: The Guilford Press.

Doabler, C. T., Cary, M. S., Jungjohann, K., Clarke, B., Fien, H., Baker, S., … & Chard, D. (2012). Enhancing core mathematics instruction for students at risk for mathematics disabilities. *Teaching Exceptional Children,* 44(40, 48-57.

*Duhon, G. J., Noell, G. H., Witt, J. C., Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2004). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33*(3), 429-443.

Eckert, T. L., Ardoin, S. P., Daly, E. J. & Martens, B. K. (2002). Improving oral reading fluency: A brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35*(3), 271-281.

*Everett, G. E., Swift, H. S., McKenney, E. L. W., & Jewell, J. D. (2016). Analyzing math-to-mastery through brief experimental analysis. *Psychology in the Schools, 53*(9), 971-983.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters,* 27(8), 861-874.

Fields, A. (2005). *Discovering Statistics Using SPSS: 2nd Edition.* Sage Publications: London.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education,* 41(2), 121-139.

Forman, S. G., Shapiro, E. D., Codding, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., Sanetti, L. M. H., & Stoiber, K. C. (2013). Implementation science and school psychology. *School Psychology Quarterly, 28*(2), 77-100.

Fuchs, D., Compton, D. L., Fuchs, L. S., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities, 44*, 339-347.

Fuchs, D., Mock, D. Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18*(3), 157-171.

Fuchs, D. & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly,* 41(1), 93-99.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next generation approach to multilevel prevention. *Exceptional Children, 78*, 263-279.

Fuchs, L. S. & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488-500.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L. & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children, 73*(3), 311-330.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239-256.

Fuchs, L. S. & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45*(3), 195-203.

Gast, D. L. & Ledford, J. R. (2014). *Single Case Research Methodology: Applications in Special Education and Behavioral Sciences. (2nd Edition).* New York: Routledge.

Geary, D. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities,* 37(1), 4-15.

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics, 32*, 250-263.

Gersten, R., Clarke, B. S., Haymond, K. & Jordan, N. C. (2011). Screening for mathematics difficulties in K-3 students. *Center on Instruction.*

*Gilbertson, D. M. (2001). *The effect of antecedent and consequent strategies on increasing student homework compliance and academic achievement.* Doctoral dissertation. Retrieved from ProQuest Dissertations.

*Gilbertson, D., Witt, J. C., Duhon, G. & Dufrene, B. (2008). Using brief assessments to select math fluency and on-task behavior interventions: An investigation of treatment utility. *Education and Treatment of Children, 31*(2), 167-181.

Haring, N. G. & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). *The fourth R: Research in the classroom.* Columbus, OH: Merrill.

Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). *The fourth R: Research in the classroom.* Columbus, OH: Merrill.

Hasbrouck, J. & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*(7), 636-644.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin,* 92, 490-499.

Hintze, J. M., Christ, T. J., & Keller, L. A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? *School Psychology Review, 31*(4), 514-528.

Hintze, J. M. & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review, 34*(3), 372-386.

*Hofstadter-Duke, K. L. & Daly, E. J. (2014). Identifying controlling variables for math computation fluency through experimental analysis: The interaction of stimulus control and reinforcing consequences. *Behavior Modification,* 1-23.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 162-179.

Howell, K.W. & Nolet, V. (1999). *Curriculum-based evaluation: Teaching and decision making (3rd ed.).* Belmont, CA: Wadsworth.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*(4), 582-600.

Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., … & Tennant, J. (2014). A critical review of five commonly used social-emotional and behavioral screeners for elementary or secondary school. *Contemporary School Psychology,* 18(4), 241-254.

Jimerson, S. R., Burns, M. K. & VanDerHeyden, A. M. (2016). From response to intervention to multi-tiered systems of support: Advances in the science and practice of assessment and intervention. In *Handbook of Response to Intervention* (pp. 1-6). Springer: Boston, MA.

Johnson, E. S., Jenkins, J. R., & Petscher, Y. (2010). Improving the accuracy of a direct route screening process. *Assessment for Effective Intervention,* 35, 131-140.

Jones, K. M. & Wickstrom, K. F. (2002). Done in sixty seconds: Further analysis of the brief assessment model for academic problems. *School Psychology Review, 31*(4), 554-568.

Kilgus, S. P., Riley-Tillman, T. C., & Kratochwill, T. R. (2016). Establishing interventions via a theory-driven single case design research cycle. *School Psychology Review, 45*(5), 477-498.

*Kleinert, W. L. (2017). *Using brief experimental analysis to evaluate the impact of combining skill and performance components in a mathematics treatment package.* Doctoral dissertation. Retrieved from PsychInfo.

Klingbeil, D. A., Van Norman, E. R., Nelson, P. M., & Birr, C. (2019). Internal likelihood ratios: Applications for gated screening in schools. *Journal of School Psychology, 76*, 107-123,

Kratochwill, T. R., Levin, J. R., & Horner, R. H. (2018). Negative results: Conceptual and methodological dimensions in single-case intervention research. *Remedial and Special Education, 39*(2), 67-76.

Lentz, F. E. (1988). Effective reading interventions in the regular classroom. In J. L. Graden, J. E. Zins, & M. L. Curtis (Eds.) *Alternative educational delivery systems: Enhancing instructional options for all students.* (pp.351-370). Washington, DC: National Association of School Psychologists.

Lembke, E. S., Hampton, D., & Beyers, S. J. (2012). Response to intervention in mathematics: Critical elements. *Psychology in the Schools,* 49(3), 257-272.

Lemons, C. J., Vaughn, S., Wexler, J., Kearns, D. M., & Sinclair, A. C. (2018). Envisioning an improved continuum of special education services for students with learning disabilities: Considering intervention intensity. *Learning Disabilities Research and Practices, 33*(3), 131-143.

Lovitt, T. C. (1978). Arithmetic. *The Fourth R: Research in the Classroom,* 127-166.

Ma, H. H. (2006). An alternative methods for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30*(5), 598-617.

Mandreker, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316.

Marston, D., Muyskens, P., Lau, M. & Canter, A. (2003). Problem-solving model for decision making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research & Practice,* 18(3), 187-200.

Martens, B. K. & Gertz, L. E. (2009). Brief experimental analysis: A decision tool for bridging the gap between research and practice. *Journal of Behavioral Education, 18*, 92-99.

McCallum, E., Skinner, C. H., Turner, H., & Saecker, L. (2006). The taped-problems intervention: Increasing multiplication fact fluency using a low-tech, classwide, time-delay intervention. *School Psychology Review, 35*(3), 419-434.

McComas, J. J. & Burns, M. K. (2009). Brief experimental analyses of academic performance: Introduction to the special series. *Journal of Behavioral Education, 18*, 1-4.

McComas, J. J., Wacker, D. P., Cooper, L. J., Asmus, J. M., Richman, D., & Stoner, B. (1996). Brief experimental analysis of stimulus prompts for accuracy responding on academic tasks in an outpatient clinic. *Journal of Applied Behavior Analysis, 29*(3), 397-401.

McKevett, N. M. & Codding, R. S. (2019). Brief experimental analysis of math interventions: A synthesis of evidence. *Assessment for Effective Intervention,* 1-11.

McMaster, K. M., Van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., … & Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences,* 22(1), 100-111.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8*(4), 283-298. WB Saunders.

Miller, S. P., & Mercer, C. D. (1993). Using data to learn concrete-semiconcrete-abstract instruction for students with math disabilities. *Learning Disabilities Research & Practice,* 8(2), 89-96.

Mong, M. D., Doggett, R. A., Mong, K. W., & Henington, C. (2012). An evaluation of math to mastery intervention package with elementary school students in a school setting. *Journal of Evidence-Based Practices for Schools, 13*(1), 61-79.

*Mong, M. D. & Mong, K. W. (2012). The utility of brief experimental analysis and extended intervention analysis in selecting effective mathematics interventions. *Journal of Behavioral Education, 21*, 99-118.

Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics. International Association for the Evaluation of Educational Achievement*. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

National Assessment of Educational Progress (NAEP). (2019). *Explore Results for the 2019 Mathematics Assessment.* Retrieved from https://www.nationsreportcard.gov/mathematics?grade=4.

National Center for Education Statistics (NCES). (2015). National Assessment of Educational Progress. Washington, DC: U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/.

Nelson, P. M., Van Norman, E. R., Klingbeil, D. A., & Parker, D. C. (2016). A comparison of methods to screen middle school students for reading and math difficulties. *School Psychology Review, 45*(3), 327-342.

Nelson, P. M., Van Norman, E. R., & VanDerHeyden, A. M. (2017). Reduce, reuse, recycle: The longitudinal value of local cut scores using state test data. *Journal of Psychoeducational Assessments, 35*(7), 683-694.

Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology, 39*(4), 335-355.

*Ota, M. (2008). *Responsiveness of elementary-aged students, with and without specific learning disabilities, to interventions for mathematics calculation.* Doctoral dissertation. Retrieved from ProQuest Dissertations.

Patton, J. R., Cronin, M. E., Bassett, D. S., & Koppel, A. E. (1997). A life skills approach to mathematics instruction: Preparing students with learning disabilities for the real-life math demands of adulthood. *Journal of Learning Disabilities,* 30(2), 178-187.

Poncy, B. C. & Skinner, C. H. (2011). Enhancing first-grade students' addition fact-fluency using classwide cover-copy-compare, a sprint, and group rewards. *Journal of Applied School Psychology, 27*(1), 1-20.

Poncy, B. C., Skinner, C. H., & Jaspers, K. E. (2007). Evaluating and comparing interventions designed to enhance math fact accuracy and fluency: Cover, copy, and compare versus taped problems. *Journal of Behavioral Education, 16*(1), 27-37.

Poncy, B. C., Skinner, C. H., & McCallum, E. (2012). A comparison of class-wide taped problems and cover, copy, and compare for enhancing mathematics fluency. *Psychology in the Schools, 49*(8), 744-755.

*Reisener, C. D., Dufrene, B. A., Clark, C. R., Olmi, D. J., & Tingstrom. D. H. (2016). Selecting effective interventions to increase math computation fluency via brief experimental analyses. *Psychology in the Schools, 53*(1), 39-57.

Renaissance Learning (2016). *Star Math: Technical Manual.* Wisconsin Rapids, WI. Available by request at research@renaissance.com.

Rhymer, K. N., Skinner, C. H., Henington, C., D'Reaux, R. A., & Sims, S. (1998). Effects of explicit timing on mathematics problem completion rates in african-american third-grade elementary students. *Journal of Applied Behavior Analysis, 31*(4), 673-677.

Rhymer, K. N., Skinner, C. H., Jackson, S., McNeill, S., Smith, T., & Jackson, B. (2002). The 1-minute explicit timing intervention: The influence of mathematics problem difficulty. *Journal of Instructional Psychology,* 29(4), 305-311.

Salvia, J., Ysseldyke, J., & Bolt, S. (2012). Assessment in special and inclusive education (12th ed.). Stamford, CT: Cengage.

Schmidt, J. D., Shanholtzer, A., Mezhoudi, N., Scherbak, B.., Kahng, S. (2014). The utility of brief experimental analysis for problem behavior maintained by escape from demands. *Education and treatment of children, 37*(2), 229-247.

Shapiro, E. S. (2011). *Academic skills problems: Direct assessment and intervention.* (4th ed). New York, NY: Guilford Press.

Shapiro, E. S., Dennis, M. S., & Fu, Q. (2015). Comparing computer adaptive and curriculum-based measures of math in progress monitoring. *School Psychology Quarterly, 30*(4), 470.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, K. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment,* 24(1), 19-35.

Scholin, S. E. & Burns, M. K. (2012). Relationship between pre-intervention data and post-intervention reading fluency and growth: A meta-analysis of assessment data for individual students. *Psychology in the Schools,* 49(4), 385-398.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* Guilford Press.

Shinn, M. R. (2004). Administration and scoring of mathematics computation curriculum-based measurement (M-CBM) and math facts probes for use with AIMSweb. *AIMSweb Training Workbook.*

Shinn, M. R. (2019). Personal communication, June 30th, 2019.

Shinn, M. R. & Bamonto, S. (1998). Advanced applications of Curriculum-Based Measurement: “Big ideas” and avoiding confusion. In M. R. Shinn (Ed.) *Advanced applications of curriculum-based measurement* (pp. 1-31). New York, NY: Guilford Press.

Shinn, M.R., Rosenfield, S., & Knutson, N. (1989). Curriculum-based assessment: A comparison of models. *School Psychology Review, 18*, 299–316.

*Silva, M. (2017). *Exploring the effect of intervention choice on academic performance and student acceptability.* Doctoral dissertation. PsychInfo.

Silva, M. R., Collier-Meek, M. A., Codding, R. S., & DeFouw, E. R. (2020). Acceptability assessment of school psychology interventions from 2005 to 2018. *Psychology in the Schools,* 57, 62-77.

Spectrum K-12 (2010). *Response to intervention adoption survey 2010.* Retrieved from http://sss.usf.edu/resources/presentations/2010/fasp_summer_inst2010/Resource_SLD/RTI/2010RTIAdoptionSurveyReport.pdf.

Soloman, B. G., Dawes, J. M., Duhon, G. J., & Poncy, B. C. (2018). False discovery rates when engaging in skill versus performance deficit analysis for academic instructional planning. *Assessment for Effective Intervention,* 1-10.

Strait, G. G., Smith, B. H., Pender, C., Malone, P. S., Roberts, J. & Hall, J. D. (2015). The reliability of randomly-generated math curriculum-based measurements. *Assessment for Effective Intervention, 40*(4), 247-253.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795-819.

Stickney, E. M., Sharp, L. B., & Kenyon, A. S. (2012). Technology-enhanced assessment of math fact automaticity: Patterns of performing for low- and typically achievement students. *Assessment for Effective Intervention, 37*(2), 84-94.

Szadokierski, I., Burns, M. K., & McComas, J. J. (2017). Predicting intervention effectiveness from reading accuracy and rate measures through the instructional hierarchy: Evidence for a skill-by-treatment interaction. *School Psychology Review, 46*(2), 190-200.

Turco, T. L. & Elliot, S. N. (1986). Students' acceptability ratings of interventions for classroom misbehaviors: A study of well-behaving and misbehaving youth. *Journal of Psychoeducational Assessment,* 4(4), 281-289.

VanDerHeyden, A. M. (2013). Universal screening may not be for everyone: Using a threshold model as a smarter way to determine risk. *School Psychology Review,* 42(4), 402-414.

VanDerHeyden, A. M., Witt, J. C., & Barnett, D. W. (2005). The emergence and possible futures of response to intervention. *Journal of Psychoeducational Assessment,* 23(4), 339-361.

VanDerHeyden, A. M. & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30*(3), 15-31.

VanDerHeyden, A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education*, *18*(1), 71-91.

VanDerHeyden, A. M. & Burns, M. K. (2018). Improving decision making in school psychology: Making a difference in the lives of students, not just a prediction about their lives. *School Psychology Review, 47*(4), 385-395.

VanDerHeyden, A. M., Broussard, C., & Burns, M. K. (2019). Classification agreement for gated screening in mathematics: Subskill mastery measurement and classwide intervention. *Assessment for Effective Intervention,* 1-11.

VanDerHeyden, A. M. & Codding, R. S. (2020). Belief-based versus evidence-based math assessment and instruction. *Communique, 48*(5).

VanDerHeyden, A. M., Codding, R. S., & Martin, R. (2017). Relative values of common screening measures in mathematics. *School Psychology Review, 46*, 65-87.

VanDerHeyden, A. M. & Witt, J. C. (2007). Best practices in can't do/won't do assessment. *Best Practices in School Psychology,* V2, 195-208.

Van Houten, R., & Thompson, C. (1976). The effects of explicit timing on math performance. *Journal of Applied Behavior Analysis, 9*(2), 227.

Van Norman, E. R., Nelson, P. M., & Klingbeil, D. A. (2017). Single measure and gated screening approaches for identifying students at-risk for academic problems: Implications for sensitivity and sensitivity. *School Psychology Quarterly, 32*(3), 405.

Van Norman, E. R., Nelson, P. M., Klingbeil, D. A., Cormier, D. C., & Lewka, A. J. (2018). Gated screening frameworks for academic concerns: The influence of redundant information on diagnostic accuracy outcomes. *Contemporary School Psychology, 23*(2), 152-162.

Vaughn, S. & Fletcher, J. M. (2012). Response to intervention with secondary school students with reading difficulties. *Journal of Learning Disabilities, 45*(3), 244-256.

Vollmer, T. R., Marcus, B. A., Ringdahl, J. E., & Roane, H. S. (1995). Progressing from brief assessments to extended experimental analyses in the evaluation of aberrant behavior. *Journal of Applied Behavior Analysis, 28*, 561-576.

Wei, X., Lenz, K. B., & Blockorby, J. (2012). Math growth trajectories of students with disabilities: Disability category, gender, racial and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education, 34*(3), 154-165.

What Works Clearinghouse Reviewer Guidance for Use with the Procedures and Standard Handbook (version 3.0). (2016).

Witt, J. C. & Elliot, S. N. (1985). Acceptability of classroom management strategies. In Kratochwill, T. R. (Ed.), *Advanced in School Psychology*, (Vol 4, pp. 251-288.). Hillsdale, NJ: Erlbaum.

Witt, J. C. & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools, 20*, 510-517.

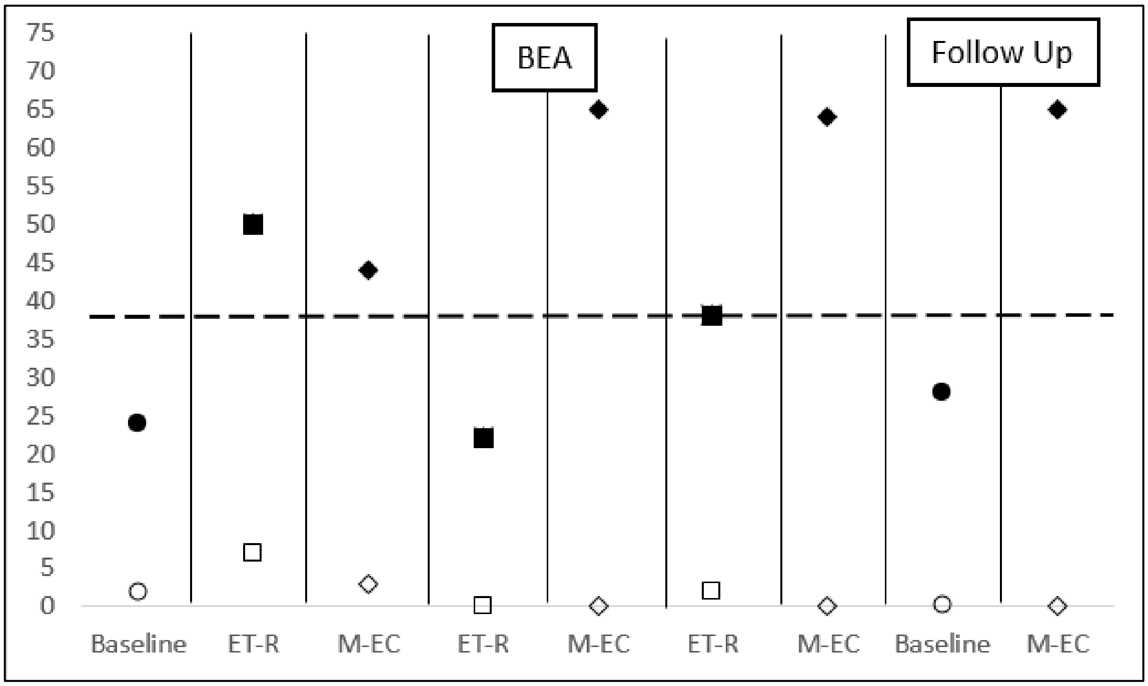
Xin, Y. P. (2008). *Quality Indicators Applied to a Single-Subject Research Study.*

Ysseldyke, J. E. & Algozzine, B. (1984). *Introduction to special education.* Boston: Houghton Mifflin.

Zigmond, N. & Miller, S. E. (1986). Assessment for Instructional Planning. *Exceptional Children, 52*(6), 501-509.

Appendix A

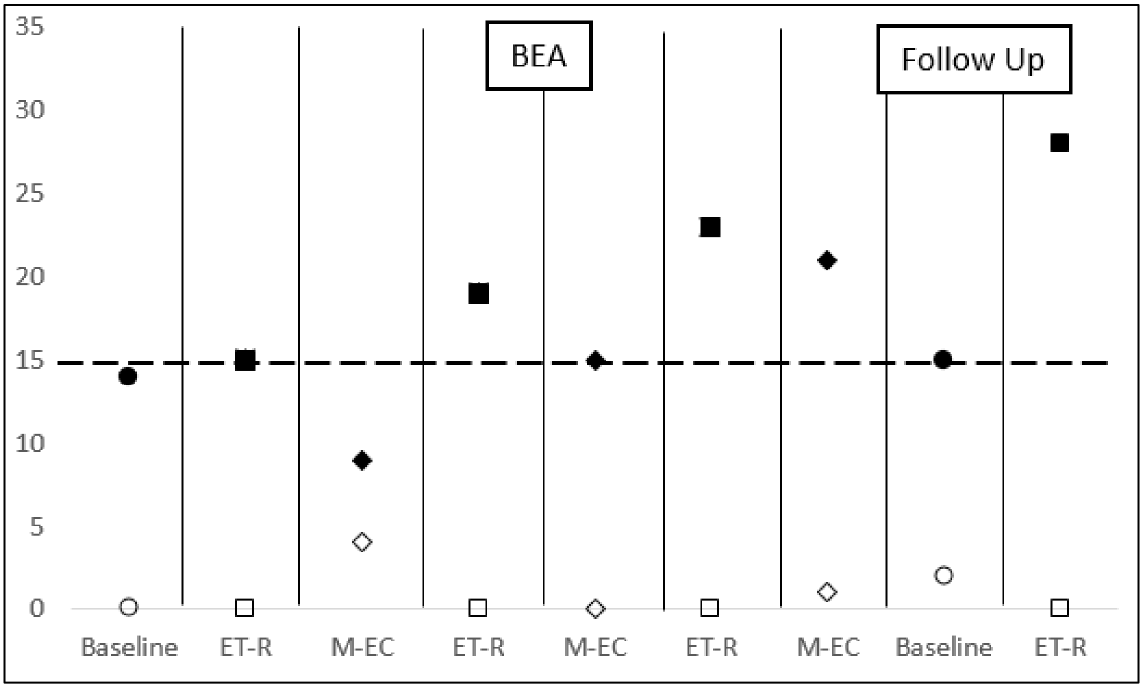
Sample Graph with All of the Effective Intervention Data Points Above the Median of the Less Effective Intervention



Note. BEA = Brief Experimental Analysis. The dotted line indicates the median of the less effective intervention.

Appendix B

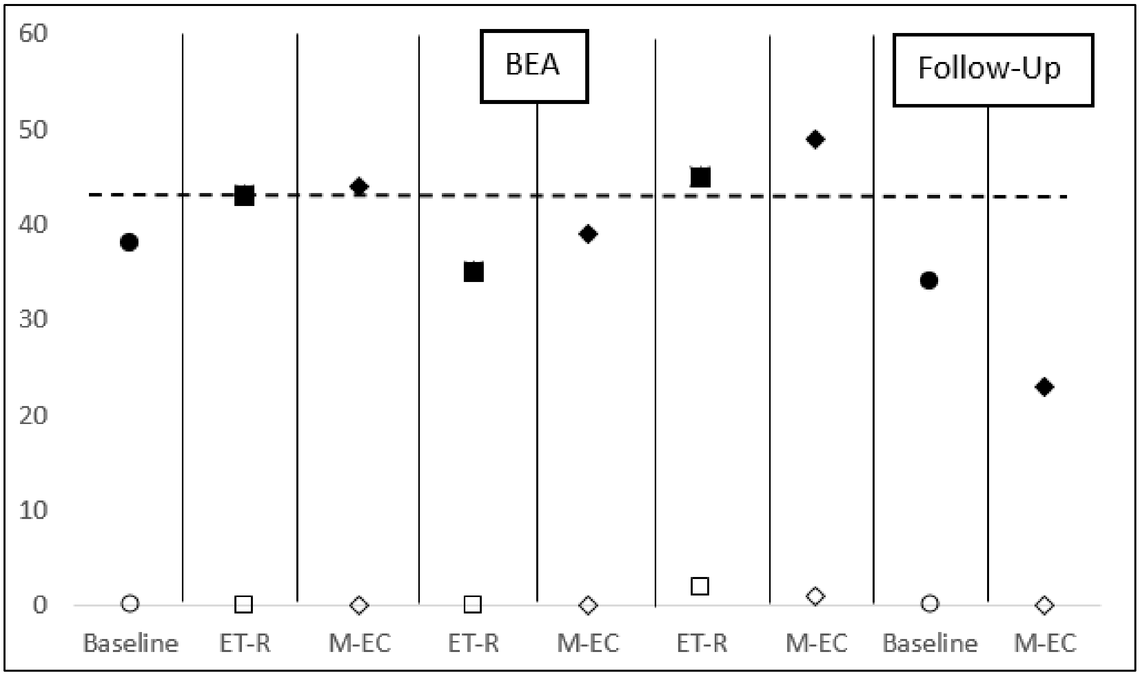
Sample Graph with 66% of the Effective Intervention Data Points Above the Median of the Less Effective Intervention



Note. BEA = Brief Experimental Analysis. The dotted line indicates the median of the less effective intervention.

Appendix C

Sample Graph with Neither Modeling with Error Correction (M-EC) or Explicit Timing with Reward (ET-R) More Effective



Note. BEA = Brief Experimental Analysis. The dotted line indicates the median of the less effective intervention.

Appendix D

Student Acceptability of Modeling with Error Correction and Explicit Timing with Reward

Item	*M*	*SD*
Modeling with Error Correction		
M-EC was fair.	5.5	1.2
M-EC was too harsh.	1.5	1.4
M-EC caused problems in my math group.	1.2	0.8
There are better ways to learn math in my classroom than M-EC.	3.7	2.1
M-EC could help other kids too.	5.4	1.5
I liked M-EC.	5.6	1.1
M-EC helped me do better in school.	5.4	1.5
Average Total Score	33.9	9.4
Explicit Timing with Reward		
ET-R was fair.	5.2	1.3
ET-R was too harsh.	2.5	2.1
ET-R caused problems in my math group.	1.7	1.5
There are better ways to learn math in my classroom than ET-R.	3.6	2.2
ET-R could help other kids too.	5.3	1.6
I liked ET-R.	5.7	1.0
ET-R helped me do better in school.	5.3	1.4
Average Total Score	32.6	9.8

Note. *n*=68. The range for each item is 1-6. 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Slightly Agree, 5= Agree, 6=Strongly Agree.

Appendix E

Teacher Acceptability of Modeling with Error Correction and Explicit Timing with Reward

Item	*M*	*SD*
Modeling with Error Correction		
M-EC would be an acceptable intervention for my students' needs.	5.2	0.6
Most teachers would find M-EC appropriate for children with similar needs.	5.0	0.6
M-EC should prove effective in supporting my students' needs.	5.1	0.5
I would suggest the use of M-EC.	4.9	0.6
My students' needs are severe enough to warrant the use of M-EC.	4.9	0.7
Most teachers would find M-EC suitable for the needs of their students.	5.1	0.5
I would be willing to use M-EC in the classroom setting.	5.1	0.5
M-EC would *not* result in negative side effects for the students.	4.7	0.7
M-EC would be appropriate for a variety of children.	5.1	0.5
M-EC is consistent with strategies I have used in classroom settings.	5.1	0.6
M-EC is a fair way to handle students' needs.	5.1	0.5
M-EC is reasonable for the needs of the child.	5.0	0.6
I like the procedures used in M-EC.	4.8	0.8
M-EC would be a good way to handle my students' needs.	5.0	0.6
Overall, M-EC would be beneficial for my students.	5.0	0.6
Average Total Score	60.8	9.8
Explicit Timing with Reward		
ET-R would be an acceptable intervention for my students' needs.	4.4	0.6
Most teachers would find ET-R appropriate for children with similar needs.	4.4	0.8
ET-R should prove effective in supporting my students' needs.	4.6	0.8
I would suggest the use of ET-R.	3.9	1.0
My students' needs are severe enough to warrant the use of ET-R.	4.1	0.9
Most teachers would find ET-R suitable for the needs of their students.	3.9	0.9
I would be willing to use ET-R in the classroom setting.	4.3	0.8
ET-R would *not* result in negative side effects for the students.	3.9	1.3
ET-R would be appropriate for a variety of children.	4.6	0.9
ET-R is consistent with strategies I have used in classroom settings.	3.8	1.4
ET-R is a fair way to handle students' needs.	3.9	0.9
ET-R is reasonable for the needs of the child.	4.2	0.8
I like the procedures used in ET-R.	3.6	1.0
ET-R would be a good way to handle my students' needs.	3.9	0.8
Overall, ET-R would be beneficial for my students.	4.1	0.9
Average Total Score	72.6	6.4

Note. *n*=14. The range for each item is 1-6. 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Slightly Agree, 5= Agree, 6=Strongly Agree.

Appendix F

Parental Consent Form

Title of Research Study: Evaluating a Skill-by-Treatment Interaction in Math Computational Fluency through Brief Experimental Analysis

Investigator: Nicole McKevett, MA, NCSP, School Psychology Doctoral Candidate

Your child is being invited to participate in a research study of how to predict which type of math instruction will work best for individual students. Your child was selected as a possible participant because your child's school agreed to participate in the current study. We ask that you read this form and ask any questions you may have before agreeing to allow your child to be in the study.

This study is being conducted by Nicole McKevett, MA, NCSP, doctoral candidate in the School Psychology program at the University of Minnesota, under the guidance of her advisor, Dr. Robin Codding, LP, BCBA, an Associate Professor at the Northeastern University, and Dr. Kristen McMaster, an Associate Professor at the University of Minnesota, and the chair of this dissertation study.

Background Information
This project is designed to understand how to best match mathematics interventions to meet the individual needs of students. Students at different stages of the learning process may benefit from different instructional strategies. The purpose of this project is to determine (a) how to best know when students are at different stages of the learning process and (b) what kinds of math instruction will work best for different stages of learning. This will allow teachers to better understand match instruction to meet students' needs.

Study Procedures
If you allow your child to participate, he or she will be asked to complete 10 math computation probes for 2-min while a researcher calculates how many digits correct per 2-min they earn. Your child will also complete a general outcome measure of computation, using a commonly using math screening measure. Then, the researcher will provide brief instruction using two different instructional procedures, which are intended to help students at different stages of learning. The researchers will then determine if it is possible to predict which intervention has the best results based on original performance on the 10 probes and the screening measure.

The study is intended to last over a one-week period for each student. If your child misses school during this week, a make-up day will occur on the following week, or as soon as your child returns to school. Below are the steps of the study that will occur:

1. The researcher will explain the study to the student, and the student will be asked if he or she would like to participate. The student will be asked to write their name on an assent form to indicate if they do/do not want to participate. If they do

not want to participate, they will be allowed to return to the classroom without any negative consequences.

2. During the first two sessions, the students will complete the first 10 math probes (five probes per session), and the general outcome math screening measure, in a large group format.

3. Over the next three sessions, the student will work one-on-one with a researcher who will provide instruction, alternating between the two mathematics interventions. In one of the interventions, the student will be offered a reward for beating their original score.

Risks and Benefits of Participation
There are minimal risks to participating in the study. However, your child may become fatigued or frustrated when completing the math tasks. If this occurs, your child will be allowed to take a break. Additionally, your child will be asked to work one-on-one with a researcher. The researcher will work to establish a relationship with the student before beginning the instruction; however, working one-on-one with an unfamiliar adult may be uncomfortable for some children. Additionally, while we will work to conduct this study during non-instructional time, there is a possibility that some class time will be missed.

There are also benefits of participation. First, information about what type of math instruction that can lead to the best outcomes in math computation will be provided to the teacher and may or may not be used in future instruction. Additionally, your child's participation will help us learn how to predict the best mathematics instruction for students, which will help future children learn to master computational skills.

Compensation
Your child will be offered a small reward during one of the interventions, which may include a sticker, pencil, fruit snacks, etc.

Confidentiality
The records of this study will be kept confidential. In any sort of report that is published or presented, we will not include any identifying information that would make it possible for anyone to identify study participants. Research records will be kept behind a locked door, and only those researchers who assist in the project will have access to these records. Any electronic data will be kept on a secured platform, which requires a dual-code entry method.

Voluntary Nature of the Study
Participation in this study is completely voluntary. Your decision to let your child participate will in no way affect you or your child's current or future relations with the University of Minnesota or their elementary school. Additionally, during this study, you or your child are free to withdraw from the study at any time, without explanation.

Contact Information
The researchers conducting this study and their contact information are below.

Nicole McKevett, MA, NCSP
Doctoral Candidate, School Psychology
weber581@umn.edu

Dr. Robin Codding, LP, BCBA
Associate Professor
r.codding@northeastern.edu

This research has been reviewed and approved by the Institutional Review Board (IRB) within the Human Research Protections Program (HRPP). If you would like to share feedback privately with the HRPP about you or your child's research experience, you can call the Research Participants' Advocate Line at 612-625-1650 or go to www.irb.umn.edu/report.htlm. You are encouraged to contact the HRPP if:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your or your child's rights as a research participant.
- You want to get additional information or provide input about this research.

Parental Permission Form

I have read the above information. Your signature documents your permission for the named child to take part in this research.

____
Printed name of child participant

____ ____________
Printed name of parent or individual legally authorized to consent for the child to participate — Date

____ ____________
Signature of parent or individual legally authorized to consent for the child to participate — Date

Appendix G

Student Assent Form

Title of Research Study: Evaluating a Skill-by-Treatment Interaction in Math Computational Fluency through Brief Experimental Analysis

You are being asked to participate in a research project. Research is a way to learn new things, by asking a question, planning, and testing it.

We are asking you to participate because you in second or third grade and you are working on learning your math facts. If it's ok with you, I'm going to ask you to work on some math problems. I'll also work with you on some math using blocks, timers, and some rewards, like pencils, stickers, and fruit snacks. The activities will look like this (show sample math probes).

If you agree to be in the study, you'll work with a researcher, like me, and practice working on math over the next week. Each time you work with the researcher will last between 10 and 25 minutes. The researcher will count how many numbers you get correct – sometimes the researcher will help you figure out the math problems, and sometimes she won't.

You might not like doing the math problems or might get tired. That's ok – we'll let you take a break when you get tired, and if you don't like being in the study, you can stop at any time. However, participating in the study will help us figure out how to help kids just like you learn to do math.

The results will just be for research. Being in the study is totally up to you, and no one will be mad at you if you don't want to do it.

Signing here means that you have read this paper, or I read it to you, and you are willing to be in the study. If you don't want to be in the study, don't sign the paper.

Remember, being in the study is up to you, and no one will be mad at you if you don't sign this, or if you change your mind later.

______________________________	_______________
Signature of child	Date
______________________________	_______________
Printed name of child	Date
______________________________	_______________
Printed name of person obtaining assent	Date

Appendix H

Baseline Instructions

The baseline for this study includes 10 math computation probes (each complete for 2-min), administered over two sessions.

Materials needed:

- 10 math computation probes (2 packets, with 5-probes each, in student folders)
- Pencils
- Stopwatch

Directions:

1. *Today we will be completing some math problems together! You will have 2-minutes to complete as many problems as you can. Make sure to try every problem. If you come to one that you don't know, mark an 'X' through it, and move on to the next one. Do you have any questions?*
   Step 1 Completed ___________

2. Say "*Begin*" and start the stopwatch when the student begins the first problem.
   Step 2 Completed ___________

3. Walk around the group and make sure that students are completing the worksheets correctly. Give encouragement to continue working if needed.
   Step 3 Completed ___________

4. After 2-min, say *"Stop. Put your pencils down and turn to the next page. You will have 2 more minutes to complete as many problems as you can. Ready? Begin."*
   Step 4 Completed ___________

5. Repeat Steps 1-3 until 5 probes have been completed.
   Step 5 Completed ___________

6. Say *"You did great work today! Thank you for your hard work!"* Allow the students to select a pencil or sticker.
   Step 6 Completed ___________

Appendix I

Explicit Timing with Reward Instructions and Fidelity Monitoring

Materials:

- ET worksheets
- Pencil
- Implementation checklist
- Rewards

Student Name: _______________

Date: ________________________

Intervention Step	Complete?	Not Applicable (Explain)
1. Pass out the ET worksheets to students and instruct them to write their names at the top of the paper.		
2. Read the following directions: **"Today we are going to complete math worksheets using explicit timing. With explicit timing, I am going to give you 1 minute to complete as many problems as you can. Your first goal is to complete each problem correctly and to not skip around. In addition, push yourself to work as quickly as possible. Are there any questions?"**		
3. Continue: **"When I say 'Begin" start answering the problems on your worksheet. Start at the top and work across the page and then go onto to the next row. Try each problem and do not skip any problems. I'm going to give you 1 minute to complete as many problems as you can. Are you ready?"**		
4. Say: **"Begin!"** Start timer for 1-min. 5. After 1 minute goes by, stop students. Record the number of problems completed correctly on the side of the worksheet. *Repeat 4 times and record the number of problems completed correctly each time.*		
6. Administer CBM for progress monitoring.		

Adapted from Poncy & Duhon (2017) Explicit Timing Intervention Packet, MIND: Facts on Fire.

Appendix J

Modeling with Error Correction Self-Monitoring Form

Materials:

- M-EC worksheets
- Pencil
- Blocks
- Number line
- Implementation checklist
- Stopwatch

Student Name: ________________

Date: ________________________

Intervention Step	**Complete?**	**Not Applicable (Explain)**
1. Pass out the M-EC worksheets to students and instruct them to write their names at the top of the paper.		
2. Read the following directions: **"Today we are going to complete math worksheets together, using blocks and a number line. First, I'm going to show you how to solve the problems. Watch me as I complete the worksheet."** Model completion of the problems on the probe for 1-min, while the student follows along.		
3. Continue: **"Now, you'll get to try to do as many of the problems as you can, for 2-min. I'm going to follow along, and we'll use the blocks and number line to work through ones that are tricky. Are you ready?"**		
4. Say: **"Begin!"** Start timer for 2-min.		
5. Provide error correction on as many errors as possible within a 5-min period. a) Model the problem using a set of blocks (e.g., "Six blocks, plus 2 blocks, equals 8 blocks.")		

b) Model the problem using the number line (e.g., "I start at 6, and count up 2…7, 8. 6 plus 2 equals 8.") c) Have the student model the problem with the blocks and/or number line. d) Solve the abstract problem on the worksheet.		
6. Administer CBM for progress monitoring.		

Appendix K

Intervention Rating Profile – 15 (IRP-15) For Explicit Timing with Reward

The purpose of this questionnaire is to obtain information on teacher's opinions of the use of **explicit timing with reward** to improve students' computational fluency (the automaticity and accuracy of retrieval of math facts). Please check the box which best describes your agreement or disagreement with each statement.

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. Explicit timing with reward would be an acceptable intervention for my students' needs.						
2. Most teachers would find explicit timing with reward appropriate for children with similar needs.						
3. Explicit timing with reward should prove effective in supporting my students' needs.						
4. I would suggest the use of explicit timing with reward to other teachers.						
5. My students' needs are severe enough to warrant the use of explicit timing with reward.						
6. Most teachers would find explicit timing with reward suitable for the needs of their						
7. I would be willing to use explicit timing with reward in the classroom setting.						

8. Explicit timing with reward would *not* result in negative side effects for the						
9. Explicit timing with reward would be appropriate for a variety of children.						
10. Explicit timing with reward is consistent with those I have used in classroom settings.						
11. Explicit timing with reward is a fair way to handle the students' needs.						
12. Explicit timing with reward is reasonable for the needs of the child.						
13. I like the procedures used in explicit timing with reward.						
14. Explicit timing with reward would be a good way to handle my students' needs.						
15. Overall, explicit timing with reward would be beneficial for my						

Source: Adapted from Witt, J.C. & Elliott, S.N. (1985). Acceptability of classroom intervention strategies. In Kratochwill, T.R. (Ed.), Advances in School Psychology, Vol. 4, 251 – 288. Mahwah, NJ: Erlbaum. Reproduced under Fair Use of copyrighted materials for education, scholarship, and research. 17 U.S.C. § 107.

Appendix L

Intervention Rating Profile – 15 (IRP-15) For Modeling with Error Correction

The purpose of this questionnaire is to obtain information on teacher's opinions of the use of **modeling with error correction** to improve students' computational fluency (the automaticity and accuracy of retrieval of math facts). Please check the box which best describes your agreement or disagreement with each statement.

	Strongly Disagree	Disagree	Slightly Disagree	Slightl y Agree	Agre e	Strongl y Agree
1. Modeling with error correction would be an acceptable intervention for my students'						
2. Most teachers would find modeling with error correction appropriate for children with similar needs.						
3. Modeling with error correction should prove effective in supporting my students' needs.						
4. I would suggest the use of modeling with error correction to other teachers.						
5. My students' needs are severe enough to warrant the use of modeling with error correction.						
6. Most teachers would find modeling with error correction suitable for the needs of their students.						

7. I would be willing to use modeling with error correction in the classroom setting.						
8. Modeling with error correction would *not* result in negative side effects for the students.						
9. Modeling with error correction would be appropriate for a variety of children.						
10. Modeling with error correction is consistent with those I have used in classroom settings.						
11. Modeling with error correction is a fair way to handle the						
12. Modeling with error correction is reasonable for the needs of the child.						
13. I like the procedures used in modeling with error correction.						
14. Modeling with error correction would be a good way to handle my students' needs.						
15. Overall, modeling with error correction would be beneficial for my students						

Source: Adapted from Witt, J.C. & Elliott, S.N. (1985). Acceptability of classroom intervention strategies. In Kratochwill, T.R. (Ed.), Advances in School Psychology, Vol. 4, 251 – 288. Mahwah, NJ: Erlbaum. Reproduced under Fair Use of copyrighted materials for education, scholarship, and research. 17 U.S.C. § 107.

Appendix M

Children's Intervention Rating Profile
Explicit Timing with Reward

Name: **Teacher:**

Grade:

Directions: I'm going to read to you some statements about the activities that we did together. Think about each one, and then let me know how much you agree with the statements. You can point on the sheet where your feelings are, with 1 being "I do not agree at all" and 6 being "I completely agree."

	I do not agree at all 1	2	3	4	5	I completely agree 6
1. Explicit timing with reward was fair.						
2. Explicit timing with reward was too harsh.						
3. Explicit timing with reward caused problems in my class.						
4. There are better ways to handle problems in my classroom						
5. Explicit timing with reward could help other kids too.						
6. I liked explicit timing with reward.						
7. Explicit timing with reward helped me do better in school.						

Adapted from Witt & Elliott, 1985

Appendix N

Children's Intervention Rating Profile
Modeling with Error Correction

Name: **Teacher:**

Grade:

Directions: I'm going to read to you some statements about the activities that we did together. Think about each one, and then let me know how much you agree with the statements. You can point on the sheet where your feelings are, with 1 being "I do not agree at all" and 6 being "I completely agree."

	I do not agree at all 1	2	3	4	5	I completely agree 6
1. Modeling with error correction was fair.						
2. Modeling with error correction was too harsh.						
3. Modeling with error correction caused problems in my class.						
4. There are better ways to handle problems in my classroom						
5. Modeling with error correction could help other kids too.						
6. I liked modeling with error correction.						
7. Modeling with error correction helped me do better in school.						

Adapted from Witt & Elliott, 1985